

2015

Climate and Weather Risk in Natural Resource Models

Nathaniel Henry Merrill
University of Rhode Island, Nateme16@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Merrill, Nathaniel Henry, "Climate and Weather Risk in Natural Resource Models" (2015). *Open Access Dissertations*. Paper 370.
https://digitalcommons.uri.edu/oa_diss/370

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

CLIMATE AND WEATHER RISK IN NATURAL RESOURCE MODELS

BY

NATHANIEL HENRY MERRILL

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ENVIRONMENTAL AND NATURAL RESOURCE ECONOMICS

UNIVERSITY OF RHODE ISLAND

2015

DOCTOR OF PHILOSOPHY DISSERTATION
OF
NATHANIEL HENRY MERRILL

APPROVED:

Thesis Committee:

Major Professor	Todd Guilfoos
	James Opaluch
	Arthur Gold

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2015

ABSTRACT

This work, consisting of three manuscripts, addresses natural resource management under risk due to variation in climate and weather. In three distinct but theoretically related applications, I quantify the role of natural resources in stabilizing economic outcomes. Manuscript 1 examines policies for controlling the risk of outbreaks of cyanobacteria, algae blooms, stimulated by excess nutrient loading transported by uncertain and possibly changing rainfall patterns for a drinking water watershed in Rhode Island. The manuscript is based on a paper co-authored with Jim Opaluch. Manuscript 2 addresses welfare outcomes of optimal, as well as open access, extraction of a groundwater resource when demand is variable based on precipitation conditions. The groundwater stock is treated as depletable both in height and spatial extent. The model is applied to the Ogallala aquifer in Kansas. The manuscript is based on a paper co-authored with Todd Guilfoos. Manuscript 3 models the market for natural gas used by the electric sector in New England as a function of weather and pipeline capacity. The effect of changes in natural gas import capacity on natural gas price and quantity consumed as a function of the severity of winter weather conditions is quantified for the New England region.

In Manuscript 1, we address policy designed to effect the risk of cyanobacteria blooms in a drinking water reservoir through watershed wide policy. Combining a hydrologic and economic model for a watershed in Rhode Island, we solve for the efficient allocation of best management practices (BMPs) on livestock pastures to meet a

monthly risk-based as well as mean-based water quality objective. In order to solve for the efficient allocations of nutrient control effort, we optimize a probabilistically constrained integer-programming problem representing the choices made on each farm and the resultant conditions that support cyanobacteria blooms. In doing so, we employ a genetic algorithm (GA). We hypothesize that management based on controlling the upper tail of the probability distribution of phosphorus loading implies different efficient management actions as compared to controlling mean loading. We find a shift to more intense effort on fewer acres when a probabilistic objective is specified with cost savings of meeting risk levels of up to 25% over mean loading based policies. Additionally, we illustrate the relative cost effectiveness of various policies designed to meet this risk-based objective.

Rainfall and the subsequent overland runoff is the source of transportation of nutrients to a receiving water body, with larger amounts of phosphorus moving in more intense rainfall events. We highlight the importance of this transportation mechanism by comparing policies under climate change scenarios, where the intensity of rainfall is projected to increase and the time series process of rainfall to change. The climate change scenarios show a shift towards a heightened risk of conditions supporting blooms and an increasing importance of spatial prioritization of nutrient control effort.

In Manuscript 2, we introduce a new economic groundwater model that incorporates the gradual shift from irrigation to dryland farming as parts of an aquifer run dry. We

accomplish this using an upside down cone to represent the spatial depletion, where the area of irrigable land above the aquifer shrinks as the water level decreases. Depletion of the aquifer may interact with uncertainty of the supply of water because the buffer that groundwater provides is no longer available. In this work, we identify the impact of spatial depletion on welfare gains from optimal management when rainfall is stochastic and follows a Markov process. Using a stylized model and dynamic programming, we estimate gains from moving away from current myopic extraction behavior to optimal use of the resource.

When applied to Kansas over a section of the Ogallala Aquifer, we find gains from management ranging from 2.88% to 3.01% with larger gains achieved under uncertainty in the rainfall process. We find that including the dynamic of the gradual spatial depletion of the aquifer does materially impact welfare results compared to other estimates of the same region. Surprisingly the serial correlation of rainfall matters little. Empirically, multi-year droughts combined with the loss of access to the aquifer only slightly increases welfare gains due to the availability of dryland farming and the productivity of that option as a backstop when available.

Manuscript 3 empirically estimates the effect of an increase in natural gas pipeline capacity in New England on monthly equilibrium natural gas prices and quantities for the electric sector. Weather plays an important role in defining the demand for natural gas due to its use for heating and electricity generation in the winter and through

electricity demand for cooling in the summer. The cost of natural gas has important consequences to the wellbeing and cost of living for millions of customers either relying directly on natural gas for heating, or electric energy consumers indirectly. This paper presents results of reduced form price and quantity time series regressions using Generalized Least Squares (GLS) followed by results of a dynamic simultaneous equation model (SEM) of the market system. Using derived empirical relationships, prices and quantities for natural gas are estimated under various weather scenarios as well as under current and expanded capacity. I highlight the role capacity has in effecting the variability of the price of energy to the region.

This work adds to the literature by providing empirical evidence and the quantification of the effect of constrained pipeline supply in an important energy market, where weather conditions, multiple demand sectors and alternative fuels determine the cost of energy. I find that capacity is a significant factor in the prices and quantities of natural gas consumed by the electric sector, with an increase in pipeline capacity of 1% leading to an average decrease in price of .48% and an increase in consumption of .2%. The SEM model finds both supply and demand to be price inelastic.

ACKNOWLEDGMENTS

I have many people to acknowledge that have taught, advised and supported me in a number of different ways. Todd Guilfoos has been a thoughtful, patient and insightful advisor over the past years. It has been a pleasure working with him on issues that we both find important and interesting, in addition to his guidance and perspective on making my way in the economics discipline. Through working with him as a research assistant, I am deeply grateful for the freedom he provided to explore problems as well as giving me the needed push to get ideas moving towards publishable work.

I would like to thank Emi Uchida for taking me in as a fresh graduate student and involving me in interesting and multi-disciplinary work early in my time at URI. I will forever be grateful for the opportunity to work with her and the rest international team in Tanzania. The experience has given me perspective and appreciation for how economics can be applied where natural resources and society are so intimately related. It was nothing short of life changing.

I owe a great deal to Jim Opaluch. He has been invaluable in guiding me through the program, particularly in formulating the dissertation. His experience, perspective and particularly his big picture insights have made my work stronger. I would not have made it to where I am without his generous references. The same can be said for Arthur Gold. In the most challenging moments of my graduate work, his kind words and targeted advice were deeply appreciated. His thoughtfulness and patience addressing my natural science inquiries opened my mind to the thinking of other

disciplines. I owe a richer understanding of the world and its people to our many conversations and the opportunity to travel to Xiamen, China.

Edson Okwelum, by being a good friend and sharp economic mind, deserves a good deal of credit for helping me think through some of the hardest parts of my work. Over many lunches and impromptu whiteboard sessions, he helped me through classwork, as well as imparting a deeper understanding of economics and its limitations. I would like to acknowledge the rest of the fellow graduate students whom I have worked with over the years, in particular, Tinting Liu, Andrew Boslett, George Sfinarolakis, Carrie Gill, Clayton Michaud, Kyle Montanio, and Brandon Elsner. Working with you all on presentations, problem sets, papers and half thought out ideas has been a pleasure. Thank you, Judy Palmer for all the support through the program.

My wife, Eleanor, has been a source of inspiration and support. I am eternally grateful for all the things, big and small, she has done to support me through this process. I cannot overstate the positive difference she has made in my life. I greatly appreciate the support from my family, Roger, Maggie, Finn and Mwanasha Merrill. I am grateful for all your love and encouragement. Thank you.

PREFACE

This dissertation is written in the three-manuscript form. The first manuscript is co-authored with Jim Opaluch. The second manuscript is co-authored with Todd Guilfoos and is under review at the Journal of Environmental Economics and Management.

Manuscript 1: Controlling Nutrient Pollution and the Risk of Cyanobacteria Blooms

Manuscript 2: Optimal Groundwater Management Under Uncertainty and Spatial Depletion

Manuscript 3: Weather, Energy and Capacity: Dynamics of the Natural Gas Market in New England

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	vi
PREFACE	viii
LIST OF TABLES	xi
LIST OF FIGURES	xii
MANUSCRIPT 1	1
Abstract	2
Introduction	3
Background	4
Conceptual Model	6
Empirical Application	12
Optimization	16
Policy Scenarios	19
Climate Scenarios	22
Results	25
Conclusion	32
Acknowledgments	34
REFERENCES	36
Figures and Tables	40
MANUSCRIPT 2	50
Abstract	51
Introduction	52
Conceptual Model	57
Empirical Model	69
Results	73
Discussion	76
Conclusion	81

REFERENCES	82
Tables and Figures.....	86
MANUSCRIPT 3	96
Abstract	97
Introduction	98
Background	99
Data	109
Model.....	111
Reduce Form Results.....	118
Increase in Capacity Scenario Analysis	120
SEM Model	123
Discussion	129
REFERENCES	134
Tables and Figures.....	138
MANUSCRIPT 2 APPENDIX	153
MANUSCRIPT 3 APPENDIX	158

LIST OF TABLES

Manuscript 1

TABLE 1: BMP CHARACTERISTICS.....	40
TABLE 2: CLIMATE CHANGE RAINFALL PARAMETERS	41

Manuscript 2

TABLE 1: PARAMETER VALUES FOR NORTHWEST KANSAS	86
TABLE 2: MARKOV CHAIN TRANSITION PROBABILITY MATRIX.....	87
TABLE 3: WELFARE GAINS FROM GROUNDWATER MANAGEMENT.....	88

Manuscript 3

TABLE 1: VARIABLE DEFINITIONS.....	138
TABLE 2: UNIT ROOT TESTS.....	139
TABLE 3: EQUILIBRIUM NATURAL GAS PRICE	140
TABLE 4: TEST OF TREATMENT OF TEMPERATURE VARIABLES.....	141
TABLE 5: EQUILIBRIUM NATURAL GAS QUANTITY	142
TABLE 6: FIRM CONTRACT QUANTITY	143
TABLE 7: EQUILIBRIUM PRICE UNDER INCREASED CAPACITY SCENARIOS	144
TABLE 8: EQUILIBRIUM QUANTITIES UNDER INCREASED CAPACITY SCENARIOS.....	145
TABLE 9: 3SLS SEM ESTIMATES.....	146
TABLE 10: 2SLS SEM ESTIMATES.....	147

LIST OF FIGURES

Manuscript 1

FIGURE 1: SCITUATE RESERVOIR WATERSHED, RI.....	42
FIGURE 2: CDF OF MONTHLY P CONCENTRATIONS AT THE RESERVOIR	43
FIGURE 3: TOTAL COSTS AT RISK LEVELS	44
FIGURE 4: DIFFERENCES IN ACRES WITH BMPs.....	45
FIGURE 5: DISCRETE DIFFERENCES IN INSTALLED BMPs	46
FIGURE 6: CDF OF MONTHLY P CONCENTRATIONS AT RESERVOIR, RCP 60	47
FIGURE 7: TOTAL COSTS AT RISK LEVELS, RCP 60	48
FIGURE 8: TOTAL COST CURVES UNDER RAINFALL SCENARIOS.....	49

Manuscript 2

FIGURE 1: SPATIAL DEPLETION MODEL	89
FIGURE 2: WATER LEVEL AND STORAGE CHANGE IN THE HIGH PLAINS AQUIFER.....	90
FIGURE 3: CROP YIELD FUNCTIONS	91
FIGURE 4: OPTIMAL VALUE FUNCTIONS	92
FIGURE 5: OPTIMAL POLICY FUNCTIONS AT GROUNDWATER HEIGHTS.....	93
FIGURE 6: GROUNDWATER HEIGHT OVER TIME	94
FIGURE 7: ESTIMATED USABLE LIFETIME OF OGALLALA IN KANSAS.....	95

Manuscript 3

FIGURE 1: NATURAL GAS CONSUMPTION BY SECTOR.....	148
FIGURE 2: ELECTRIC ENERGY GENERATION BY FUEL SOURCE	149
FIGURE 3: NATURAL GAS PRICE BY CONSUMING SECTOR	150
FIGURE 4: LOG TRANSFORMATION Q-Q PLOT TEST.....	151
FIGURE 5: SIMULATION OF EQUILIBRIUM PRICE AND QUANTITY	152

Manuscript 1

Controlling Nutrient Pollution and the Risk of Cyanobacteria Blooms

By

Nathaniel H. Merrill, Jim Opaluch

Department of Environmental and Natural Resource Economics

University of Rhode Island

To be submitted to American Journal of Agricultural Economics

Controlling Nutrient Pollution and the Risk of Cyanobacteria Blooms

Nathaniel H. Merrill, Jim Opaluch

Corresponding Author:

Nathaniel H. Merrill
Graduate Student
Department of Environmental and Natural Resource Economics
University of Rhode Island
Coastal Institute
1 Greenhouse Rd, Kingston RI 02881

Phone: (508) 878-1809; Email: nateme16@my.uri.edu

Abstract

This paper assesses policies for controlling the risk of cyanobacteria (blue-green algae) blooms stimulated by excess nutrients. We examine cost-effective strategies that are based on addressing the probability mass in the upper tail of the distribution of phosphorus loads and resultant concentrations, which are disproportionately influential in supporting episodic blooms. Combining a hydrologic and economic model for a drinking water reservoir watershed in Rhode Island, we solve for the efficient allocation of best management practices (BMPs) on livestock pastures to meet a monthly risk-based objective as well as a mean-based water quality objective. We hypothesize that management designed to control the upper tail of the probability distribution of phosphorus concentrations implies different efficient management actions as compared to controlling mean conditions. We find a shift to more intense effort on fewer acres when a risk-based objective is specified with cost savings of meeting risk levels of up to 25% over mean-based policies. Additionally, we illustrate the relative cost effectiveness of various policies designed to meet this risk-based objective.

Introduction

Water quality impairments often present in discrete events, for example: fish kills, algal blooms, red tides, and beach closures. For the cases involving excess nutrient loading, these perceivable damaging occurrences are the more noticeable result of complex interactions across the landscape between humans, climate and ecological systems. Rainfall and subsequent runoff is the transportation mechanism for nutrients from the landscape to water bodies. Since rainfall is variable, any sought after water quality benefits of policy interventions are stochastic. Changes in policy create spatial and temporal configurations of human activity that in turn affect the distribution of conditions in a water body (Milon 1987).

This paper assesses policies for controlling the risk of cyanobacteria (blue-green algae) blooms stimulated by excess nutrients. We examine cost-effective strategies that are based on addressing the probability mass in the upper tail of the distribution of phosphorus concentrations in a water body, which are disproportionately influential in supporting episodic blooms. Combining a hydrologic and economic model for a drinking water reservoir watershed in Rhode Island, we solve for the efficient allocation of best management practices (BMPs) on livestock pastures to meet a monthly risk-based as well as a mean-based water quality objective. We hypothesize that management based on controlling the upper tail of the probability distribution of phosphorus (P) concentrations in the reservoir implies different efficient management actions as compared to controlling mean conditions. We find a shift to more intense effort on fewer acres when a risk-based objective is specified with cost savings of meeting risk levels of up to 25% over mean-based policies. Additionally, we illustrate

the relative cost effectiveness of various policies designed to meet this risk-based objective.

Rainfall and the subsequent overland runoff is a source of transportation of nutrients to a water body, with larger amounts of P moving in more intense rainfall events. We highlight the importance of this transportation mechanism by comparing policies under climate change scenarios, where the intensity of rainfall is projected to increase and the time series process of rainfall is projected to change for Rhode Island. The climate change scenarios show a shift towards a heightened risk of conditions supporting blooms and an increasing importance of spatial prioritization of nutrient control effort.

Background

Cyanobacteria blooms are challenging to address because they are episodic. They are patchy and difficult to predict (Hudnell 2008) because they depend upon the confluence of factors including sunlight, nutrients (especially phosphorus), pH, precipitation, water temperature, water flow, and water column stability (EPA 2012). Under favorable conditions, cyanobacteria populations can rapidly proliferate and create extensive blooms dominated by a single (or a few) species that release toxins harmful to ecosystems and humans. Even in the absence of toxic effects, cyanobacteria blooms can result in adverse effects on taste and odor of drinking water (EPA 2012).

Of the conditions supporting blooms in fresh water, P loads and resultant in-lake concentrations are the most dependent upon anthropogenic sources and most amenable

to management. P loads are stochastic since transport to receiving waters is mediated primarily by eroded soil moving in surface runoff (James & Alexander 1998; AL Heathwaite 2000; Morgan 2001). As a consequence, extreme precipitation events are influential in determining transport of P to water bodies, and contributing to conditions favorable to blooms. Policy within this stochastic context must be effective in controlling P loading in larger precipitation events that lie in the tails of the probability distribution. This could be of increasing concern in coming decades, as there are projections that climate change will increase high intensity precipitation events (IPCC 2007; Meehl et al. 2005; IPCC 2011; Kunkel et al. 2013).

The location and mix of nutrient control effort across a watershed affects both the average emissions as well as the shape of the distribution of emissions throughout varying weather conditions. When damages from P emissions are attributed to threshold based events, those damages are not a linear function of P concentrations, or continuous estimates of damages are unknown (Lathrop et al. 1998), then optimal nutrient control efforts that address mean conditions may differ greatly from those that address damaging occurrences or threshold violations (Shortle & Horan 2001). Therefore, simple rules for improving abatement efficiency by focusing on the lowest marginal cost abaters in terms of average emissions do not necessarily hold (Shortle 1990). Rather, policy must consider the upper tail of the distribution of emissions and correlations across emitters.

Policy prescriptions to address a non-point source nutrient control problem in both a static as well as dynamic setting focus on location (and possibly time) differentiated taxes, subsidies, or standards based on estimated marginal damage of runoff, ambient

conditions or input factors to approach optimal configurations of production and abatement effort (Goetz & Zilberman 2000; Segerson 1988; Shortle & Horan 2013). Many studies use mean Nitrogen and Phosphorus concentrations or reduction targets as constraints to solve for the optimal allocation of nutrient control effort (Rabotyagov et al. 2010; Gitau et al. 2004; Arabi et al. 2008; Kling 2011). A number have explored costs in addressing probabilistic objectives of nutrient concentrations (Kampas & White 2004; Elofsson 2003; Shortle & Horan 2001; Gren et al. 2002; Milon 1987; Kataria et al. 2010; Zhu et al. 1994; Huang et al. 2012) and find distributional assumptions of emissions and the strictness of the probabilistic objective effect the efficiency of nutrient control design.

This work contributes to the literature by assessing the efficiency of the design of water quality improving policy to meet a risk-based objective. In contrast to previous work, we quantify the efficiency of policy designed to address a probabilistic constraint representing the risk-based objective as compared to mean-based policy using a hydrologic and numerical optimization model, while maintaining the important discrete nature of the problem. The resultant differences between efficient and second best allocations to address a risk-based objective show not just the total cost savings, but also the practical implications for how the allocation would differ across the watershed. By integrating climate model projections, we show that climate change may shift the risk profile and increase the importance of addressing the risk-based goal directly. The results underscore the need of spatially prioritizing nutrient control activities.

Conceptual Model

We assume a watershed partly consisting of heterogeneous agricultural fields with varying physical properties (e.g., slope, soil type, location). P accumulates in the soil over time and is transported off fields into waterways with precipitation events. P flows with water downstream to a water body, henceforth, reservoir, which provides ecosystem services valued by society.

We formulate the problem as finding the least-cost allocations of nutrient control effort to meet probabilistic constraints. This is in contrast to finding an economically optimal solution, which would include a damage function explicitly into the objective function. A damage function might take many forms depending on the water quality impairment. For instance, if damages were linear in concentrations, addressing mean concentrations would be sufficient. However, if the damages were convex to concentrations, or based on a threshold where damages are minimal and increase after a particular point, then higher order moments in the distribution of conditions matter to policy design. A probabilistic constraint can handle this type of least-cost problem by addressing the shape of the distribution of conditions (Shortle 1990). Our application to cyanobacteria blooms, which are known to be particularly reliant on high levels of P concentrations, fits this type problem. The damages are uncertain, but damaging blooms are, at the least, nonlinearly related to nutrient concentrations.

The distribution of total P loading at a water body is the sum of individual sources of emission less loss in transport. Stricter correlation between the sources leads to more burdensome probabilistic constraints and higher costs (Kampas & White 2004; Elofsson 2003). Given that each farm's emissions are highly correlated when rainfall

is assumed consistent across the space, highly correlated loadings seem like a reasonable assumption, especially in a geographically small watershed. Each heterogeneous source will contribute a different but correlated distribution of nutrient loading based on its soil type, slope, inputs and location, or more generally a source's place on the P transfer continuum (Haygarth et al. 2005). We assume the watershed of concern is sufficiently small that rainfall is perfectly correlated across fields. Assumptions of the distribution of pollutant load and transport can have large implications for estimates of efficient abatement (Gren et al. 2002; Kataria et al. 2010). In the numerical application below, we rely on a physical process-based hydrologic model to estimate the important relationships.

The goal of policy is to control the risk of cyanobacteria blooms by taking actions across the watershed to maintain P concentrations in a water body, a reservoir in our application, below a target concentration to avoid conditions that are favorable for blooms. However, P concentrations are stochastic, since loads depend upon random precipitation, especially upon extreme precipitation events. As a consequence, we specify policies that control the probability that a threshold is violated. Hence, the optimization problem is formulated as finding the least-costly set of control policies such that the probability of exceeding a threshold for P concentrations in the reservoir is less than a target level, α .

Soil P concentration on a field i at time t , P_{it} , is a state variable that evolves over time following the dynamic equation

$$\Delta P_{it} = \gamma F_{it} + \beta A_{it} - \delta_i P_{it} - B_{it}(k_i(r_t, P_{it}))$$

where F represents fertilizer inputs, A represents density of animal units, γ and β represent the associated P application rates per unit, δ represents the loss rate for phosphorus (e.g., loss through plant uptake). $k_i(r_t, P_{it})$ represents the field-specific rate of P emissions associated with a precipitation rate, r_t , and soil P concentrations, P_{it} , at time t . Therefore, $k_i(r_t, P_{it})$ represents the rate of P loss associated with soil erosion and runoff from rainfall. It is commonly assumed that the relationship between soil erosion, a major source of P loading, and rainfall intensity follows a power law. This assumption is based on the characteristics of sediment movement as a result of rainfall kinetic energy (van Dijk et al. 2002), and motivates the non-linear relationship between rainfall intensity and P loading. However, erosion and nutrient loss are a function of a number of other factors including soil saturation, slope and vegetative cover. In the numerical application below, we rely on a physical process-based hydrologic model to estimate the important relationships.

P control practices (BMPs) are represented as affecting runoff and are represented by the function B_{it} . The simplest, and widely used formulation for P control is to use BMP efficiencies, which are a fixed percentage reduction of nutrient runoff based on edge of field studies. This is the assumption made in the empirical model presented below, but we will leave the theoretical model more general. Therefore, the P emissions, e_{it} , in total loading (kg) terms from one field in time t are:

$$e_{it}(k_i, P_{it}, r_t, B_{it}, s_i) = s_i B_{it}(k_i(r_t, P_{it}))$$

where s_i is the size of field i . Since the objective of the study is the water quality at the reservoir, the total loading is the sum of the loading from the n heterogeneous fields,

as well as the additional loading from all other land use types that are not in the choice set of nutrient control. Because water remains in the reservoir for a period of time (residence time), the reservoir is itself a sink of nutrients and has its own in-lake processes. The P concentration, TP_t , in the reservoir at any one point in time is a stock variable progressing by an equation of motion.

$$\Delta TP_t = L \left(\frac{\sum_{i=1}^n e_{it}(k_i, P_{it}, r_t, B_{it}, s_i) + z_t(r_t)}{V_t(r_t)}, TP_t \right)$$

ΔTP_t is a function of both aggregate loading as well as the associated water volume, $V_t(r_t)$ in order to convert the P loads (in kg) to concentrations (ppb). $z_t(r_t)$ represents the emission from all other land not considered for nutrient control actions. The function L represents the in-lake processes that might affect the progress of P concentrations.

Therefore, the stochastic optimization problem can be described as follows:

$$\begin{aligned} \min_{B_{it}} \text{COST} &= \sum_{t=1}^T \sum_{i=1}^n C_{it}(B_{it}, s_i) \\ \text{st.} \\ \Delta P_{it} &= \gamma F_{it} + \beta A_{it} - \delta_i P_{it} - B_{it}(k_i(r_t, P_{it})) \\ \Delta TP_t &= L \left(\frac{\sum_{i=1}^n e_{it}(k_{it}, P_{it}, r_t, B_{it}, s_i) + z_t(r_t)}{V_t(r_t)}, TP_t \right) \\ Pr\{\overline{TP} \leq TP_t\} &\leq \alpha \\ B_{it} &\subseteq BMP \end{aligned}$$

Where:

B_{it} - Combination of BMPs on field i in time t

BMP - Set of BMP options
 $C_{it}(B_{it}, s_i)$ - Cost of BMP employed on field i at time t
 s_i - Size of field i
 P_{it} - P soil concentration on field i in time t
 k_i - Field specific rainfall to emissions function
 F_{it} - Fertilizer applied on field i in time t
 A_{it} - Animal density on field i in time t
 δ_i - P uptake and stabilization on field
 TP_t - P concentration in reservoir
 \overline{TP} - P concentration threshold
 $e_i(r_t, P_{it}, B_{it})$ - P runoff of field i in time t
 L - In-lake functions
 $z_t(r_t)$ - P loading from land uses not in the choice space
 α - Probability of exceeding threshold (risk level)
 r_t - Rainfall

B_{it} represents the set of chosen controls given the available options (BMP), with costs in a given period $C_{it}(B_{it}, s_i)$. The choice of controls effects P runoff, e_{it} , from each heterogeneous pasture land i with environmental factors (e.g., field size, location, slope, soil type) whose runoff to rainfall relationship is defined by k_i . Each e_i distribution varies by environmental factors and its shape is determined by choices of P control on each farm. The P threshold, $\overline{TP}(r_t)$, is the concentration of P that supports damaging cyanobacteria blooms. The term $Pr\{\overline{TP} \leq TP_t\} \leq \alpha$, represents the probabilistic constraint on the reservoir concentrations being above this threshold only $\alpha\%$ of the time or less. α captures the risk-based water quality objective.

Replacing the probabilistic constraint above with the following represents the mean-based water quality objective:

$$\frac{1}{T} \sum_{t=1}^T TP_t \leq \eta$$

where η captures strictness of the mean-based water quality objective.

Empirical Application

The framework above is applied to the problem of controlling conditions supporting cyanobacteria blooms in the Scituate Reservoir watershed in Rhode Island. The Scituate Reservoir Watershed is 86 square miles and is made up of five sub-watersheds, which drain into the Scituate Reservoir system, which is owned and managed by the Providence Water Supply Board, Figure 1. The watershed is rural and mostly forested (79%) with light and medium residential (7%) and (6%) agriculture, including pasture, row crop and orchard land. Providence Water owns and manages 25% of the total watershed area as conservation land. The reservoir system provides water services to 60% of residents and business of Rhode Island including the municipalities of Providence, N. Providence, Warwick, E. Providence, Western Cranston, Kent Co., East Smithfield, Smithfield, Greenville and Lincoln accounting for 92 million gallons a day. Taken as a whole system, the water quality is good. As described by loadings of nitrogen, it is considered low risk. Phosphorus loading to surface water (.18 kg/ac/year) is only slightly elevated above natural background values (.09 kg/ac/year) (Scituate Reservoir Source Water Assessment 2003).

The Scituate Reservoir system is made up of five sub-reservoirs (Figure 1). More recently, a limnological study was conducted for the Regulating¹ and Moswansicut sub-reservoirs and found total P levels indicating a body of water in a mesotrophic to

¹ “Regulating” is a proper name for the sub-reservoir of the Scituate Reservoir system.

eutrophic state. Given the variable nature of rainfall, P concentration values ranged from below detection levels to 36 ppb, indicating a lake in a eutrophic state (ESS Group, Inc. 2011). Pollution abatement efforts is well described by the theoretical model presented earlier, since the Regulating reservoir is observed crossing a threshold of conditions for which algae blooms are present and since damages are related to discrete bloom events. A water manager may only care about average condition of the lake inasmuch as it controls the probability of exceeding the eutrophic threshold and exposure to a risk of an algae bloom. More efficient use of resources may be to target areas that have loadings, which are particularly sensitive (convex) to larger rainfall events, based on environmental factors, therefore limiting the total loading during the more extreme events. The possible resource savings and practical implications are quantified in the results.

The RI Department of Environmental Management's maximum total P standard for freshwater lakes and reservoirs is 25 ppb (Rhode Island Office of Water Resources 2009), which corresponds to the eutrophic threshold of 50 on the trophic state index (TSI) (Carlson 1977). Based on the standard Vollenweider model, a lake or reservoir with characteristics of the Regulating Reservoir with a total volume of 1314 acre feet, a mean depth of 1.7 meters and an average residence time of 21.4 days, a yearly loading of roughly 790 kg of P would correspond to an annual average P concentration of 25 ppb (Vollenweider 1971).

A water body with a concentration exceeding 50 ppb, or 60 on the TSI scale, is at a level where algae blooms occur and where cyanobacteria may dominate. This would correspond to a yearly loading of roughly 1000 kg of P. However, this yearly concentration is made up of a series of monthly loadings and water flows. Therefore to construct the monthly risk measure for this paper and with the relatively short residence time of the reservoir, we use a monthly (May-October) inflow concentration above this 50 ppb threshold as the proxy for conditions that could support cyanobacteria blooms. This inflow concentration is calculated by using the variable monthly flow into the reservoir and the monthly total P loadings. The probability that the threshold is violated is the risk-level, α in the conceptual model. This threshold and the cumulative distribution function of monthly concentrations modeled for the Regulating Reservoir can be found in Figure 2.

We use a flow-calibrated model, the Soil and Water Assessment Tool (SWAT), to estimate phosphorus loading to the reservoir as a function of rainfall and land use (Gassman et al. 2007). SWAT is a process-based water-modeling tool designed to simulate the surface and ground water quality and quantity, and to predict effects of land use change, land management practices, and climate change in complex watersheds. SWAT is a public domain model available free for download (<http://swat.tamu.edu/>). The SWAT model was developed and refined by the USDA Agricultural Research Service and Texas A&M University over a period of more than 30 years and has supported hundreds of scientific papers.

As part of a larger ecosystem service study in the watershed, our work focuses on practices installed on small acreage livestock farms to address nutrient loading and water quality in the watershed. Including variability in the animal density would be preferable as this information also describes the rainfall to P runoff relationships. However due to privacy concerns, we were not able to obtain site-specific animal counts. As a consequence, we use the average animal density in the watershed (.5 AU/acre), as described by a livestock tally for the towns overlapping the watershed conducted by the Northern Rhode Island Conservation District (NRICD). This density was applied to all pastureland in the model. Given these assumptions, loadings from pasture land (about 3% of the watershed area) represent about 7% of the total P loadings to the Regulating reservoir.

In practice, nutrient control policies can only target a subset of loading to a given water body. There are natural loadings as well as other anthropogenic sources that may not be considered for management options. Therefore, policy in any watershed affects a subset of the total loadings and may have a limited effect on water quality. However, certain loading sources may be more sensitive to rainfall intensity and runoff and nutrient control effort on this land may affect the distribution of conditions in a water body. Pasture land in particular is a large source of sediment bound P loading and a natural place to direct effort (Harmel et al. 2006). While representing a fraction of the average loadings, management actions taken may affect the water quality in the reservoir if damages are convex or based on threshold concentration. In our case, management actions could affect the probability that the eutrophic threshold is violated.

We use historic daily rainfall data collected by the Kingston Weather Station in Kingston, RI as well as output from downscaled climate models with the calibrated hydrologic model to estimate P loads and concentrations in the reservoir. SWAT is used to construct a response surface for exploring the P loading by month under different combinations of BMPs applied to pasture land. SWAT is only run once per climate rainfall scenario to generate location, soil, and land use specific loadings and water flow.

BMPs were simulated using efficiencies from the literature following common practice, as presented in Table 1. These efficiencies are estimated by edge of field studies and represent the percent of baseline loadings that are reduced by the installation of a certain practice. The effects of BMPs are assumed multiplicative when combined (Rao et al. 2009; Gitau et al. 2004). The costs of installing and operating each BMP over 20 years is based on the accounting for practices in RI in the NRCS EQUIP Payment Practice Schedules for 2012 assuming a 5% discount rate (USDA Natural Resource Conservation Service 2012). BMPs are treated as discrete practices. With three practices, this results in a total of 8 combinations on each farm. Each practice was characterized by a fixed and variable (per acreage) cost. The manure management BMP was translated from volume based variable cost to acreage based on average animal density in the watershed. These costs are shown in Table 1. For the simulation, we assume that it is a one-time decision to install and maintain the BMPs over their lifetimes.

Optimization

The objective of the numerical optimization is to solve for the least-cost allocation of BMPs across the pasture land at each risk level within the feasible set of solutions for each climate rainfall scenario. Given the set of farms and choices there are eight combinations of practices on 113 parcels resulting in a problem with 8^{113} unique combinations of practices across the watershed. These discrete choices affect the distribution of runoff from each parcel of pasture land, which, when combined with loading from all other land types, sum to the total loading and distribution of P concentrations at the reservoir. An optimization of this type is a chance-constrained integer-programming problem. The problem is large, stochastic, and non-smooth. We use Matlab 2013 and the Global Optimization Toolbox's Genetic Algorithm (GA) integer optimization routine combined with output from SWAT and the cost and BMP effectiveness estimates described above to estimate the least-cost frontier.

GAs belong to a family of evolutionary algorithms, where possible solutions (individuals) to the objective function progress through populations and generations based on their relative fitness scores representing the optimization goal. Each new generation is created by crossing over (randomly combining) elements of better-fit individuals to create offspring. Through repetition, mutation and natural selection, the population converges to a solution. GAs have been used in nutrient abatement optimization frameworks by many authors (Gitau et al. 2004; Rabotyagov et al. 2010; Jha et al. 2009; Srivastava et al. 2002; Arabi et al. 2008). In our case, the solution is the allocation of BMPs across the landscape and its corresponding costs and risk level. One of the inequality constraints is the risk level on concentrations of inflow to the

reservoir. A GA is well suited for this type of spatial problem, as it does not require linearity, continuity, or differentiability in the objective or constraint functions (Arabi et al. 2008)

The makeup of the initial population affects the efficiency of the routine as well as the width of the solution space that it searches, which is important in confirming a found solution is a global and not local optima. Taking advantage of some prior knowledge, the initial population for the GA was seeded with individuals representing uniform adoption of each practice or across the board adoption of all possible BMP combinations, as well as allocations found by incentive based policies described below. The remaining individuals were generated randomly using a uniform distribution. In order to cover a wide range of solutions, the algorithm was run with a population of 500 individuals, for a maximum of 500 generations, or until the variability of the fitness values of individuals within the population was within a function tolerance.

We solved for the least-cost allocation of BMPs for each level of risk within the lower and upper bounds of our choice set, which are the risk levels between no installations and all practices installed on every farm. We proceed by seeding the initial populations for the GA with previously found optimal solutions. We execute this process under the probabilistic constrain as well as under the mean constraint for each rainfall scenario. As a robustness check and to confirm the GA was behaving correctly, we compared the allocations found by the GA under the mean concentration

constraint (without the allocations resulting from policy scenarios in the initial population) to the simple method of sorting the farms by the average per unit cost of abatement, which defines the efficient progression for the mean concentration goal. The two solutions were identical.

Policy Scenarios

We construct cost functions for each of the following policy designs, where the function for each design indicates the cost of achieving each risk level. We compare the efficient allocation found using the optimization process above to three simulated policies: a location differentiated subsidy; a flat subsidy based on the relative effectiveness of the BMPs, but devoid of location specific factors; and a uniform standards approach. The subsidies are simulated assuming each farmer chooses the practice(s) on his/her farm to maximize the value of a one-time up front subsidy minus the discounted present cost of installing and operating the practices. We assess efficiency in terms of the cost of the installed practices to meet risk levels. The subsidy is treated as a transfer payment rather than a social cost.

The location-specific subsidy assumes a one time up-front payment is offered for installing practices, with the total payment based on the average annual amount of P abated at the reservoir due to the practice or combination of practices on a farm. The payment is fixed across the watershed for a kg of P abated, but the abatement effects of BMPs vary for a particular practice on a farm. For example, if farmer x installs a practice which would result in a average reduction of 1 kg P/year from reaching the reservoir and the watershed wide per kg P incentive is \$1000, the farmer weighs the

total one time payment of \$1000 against the cost of installing and operating that practice over 20 years.² The farmer chooses the combination of practices that maximizes net profit, including the option of no BMP installations. Clearly, this type of subsidy requires an agency to model and offer site-specific subsidy schedules based on a particular farm's heterogeneous effect, which is a tall task, both in terms of transparency, politics and operational constraints. The benefits, quantified in our results, would be a more cost-effective allocation of effort due to using the site-specific factors.

This setup is similar to the USDA's Conservation Reserve Program (CRP), where land is retired and put into environmentally friendly management practices in return for a rental rate from the government. The program relies on bids from the farmers that are then ranked by an Environmental Benefits Index (EBI). The EBI balances many environmental goals such as soil conservation, nutrient retention and wildlife habitat. The EBI is spatially differentiated by distance to water bodies, soil type and slope, as well as giving weight to certain land based on particular initiatives. The government picks the top ranked bids based on cost-effectiveness and budgetary constraints. If accepted, the yearly payments are set at the bid price the farmer submitted for the life of the contract (USDA 2015). The difference compared to the first location-specific subsidy policy is that we assume a payment is offered based on the farm's effect and the farmers choose if that is in their best interest. The equivalent in the CRP framework would be if the government offered a fixed payment based on a unit of EBI and informed a farmer of their possible payment for installing a practice on their

² An alternative way to think about this on time upfront payment would be to consider an equivalent yearly payments whose present value is equal to this \$1000.

particular farm based on that farm's total EBI. Thus, there is a difference in market mechanism, with the CRP program based on accepting asking prices from farmers, and the subsidy described in this paper based on farmers accepting bid prices from the government. These differences could have implications to participation, size of transfer payments, but not necessarily allocate efficiency in terms of social cost.

The second policy scenario, a flat subsidy, assumes a subsidy based only on information of farm size and practice, but no other site-specific information. Therefore, it is a subsidy based on an assumed per acre relative effectiveness of the BMPs and the farm size. Hence, if the payment for a practice were \$100 per acre, a practice that is twice as effective in general would receive \$200 per acre regardless of the particular farm's P loading. The installation decisions follow the same logic as above, with the farmers choosing the practices that maximize their net profit.

These subsidy based policy designs are simulated for the full range of achievable allocations by varying the payments to cover the whole range of effects. This would be the per kg P payments that would induce the range from the minimum to where maximum BMPs are installed based on the implementation choices by farmers. The allocations are evaluated based on achieving risk levels at various costs in the results section.

The final sets of policies are uniform standards. Under the uniform standard policy, each farm is required to adopt a particular combination of the three BMPs. A cost function for uniform standards is constructed by calculating the costs and associated risk levels for each of the eight possible BMP combinations.

Climate Scenarios

As indicated above, extremes in precipitation intensity are central for determining P loads. This could have important implications for P concentrations and associated risk of cyanobacteria blooms because, while significant uncertainty is involved, there is an expectation that the Northeastern US will experience an increase in total annual rainfall and an increased percent that falls in extreme precipitation events (Karl et al. 2009; IPCC 2011). We use the process described below to model rainfall parameters in order to assess management actions under baseline and future climate change scenarios, where the probability distribution and time series nature of rainfall changes.

Baseline downscaled climate model runs did not sufficiently reproduce the same statistical properties of the observed time series.³ Consequently we retained the statistical parameters of the observed series and apply the implied differences in those parameters from the climate model runs to create the future scenarios in order to make a proper comparison. The changes between the present and future runs for each of the statistical parameters for each month were calculated in percent terms. These deltas between the modeled past and future from the climate model runs were averaged by RCP scenario and applied to the historic observed parameters (1961-2010) to generate the future climate scenarios for input to the hydrologic model (Wilks 2012).

³ For practitioners, Auffhammer et al. (2013) provides a clear overview the pitfalls of using climate model output in economic analyses. When comparing the past and future projections, the same spatial scale should be used. Since the SWAT model was calibrated using the observed time series from one gauge station, and the climate model data was at 1/8°, which reproduces the properties of the average conditions in the time step (daily) of an area about 54 square miles, the modeled distribution of daily rainfall differs with the past observations.

We model precipitation with a two-stage process. First, we determine whether a day is wet or dry. Then, we determine precipitation amounts for wet days. The probability of a wet day (more than trace amounts of precipitation) is modeled using a first-order Markov process (Srikanthan & McMahon 1999). Hence, the first stage is to model the probabilities of rainfall on date t for the cases where date $t-1$ was wet and where date $t-1$ is dry. To do so, we specify the conditional probability of date t being wet when date $t-1$ was wet ($P(w_t|w_{t-1})$) and conditional probability of date t being wet when date $t-1$ was dry ($P(w_t|d_{t-1})$). We model rainfall amounts for wet days using a two-parameter gamma distribution following the methods described in Davison et al, (2005) and Geng et al. (1986). We estimate the parameters of the gamma distribution of rainfall applying method of moments matching to historic daily rainfall data and the climate model runs.

We use historic daily rainfall data for Rhode Island from 1961 to 2010 from the Kingston, RI weather station to calculate these two conditional probabilities and gamma parameters for each month for the baseline rainfall scenario. We then create a 100-year time series of simulated daily rainfall with the observed statistical properties for input to the hydrologic model. We use a longer series for the simulation than the 50-year windows used to calculate rainfall parameters in order to better populate the discrete distribution of monthly concentrations under each scenario. To create the future rainfall scenarios, we calculated the parameters described above for both historic (1950-1999) and future (2050-2099) runs of 55 statistically downscaled

climate model runs from the CMIP5 multi-model ensemble⁴ made available at the Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections archive (Reclamation 2013). Model runs were from both the RCP 6.0 (representative concentration pathways) and RCP 8.5 emission scenarios (Reclamation 2013). Different teams of researchers developed each pathway with alternate assumptions on technology and economic growth. Descriptions of these can be found at the RCP Database (2014). The RCP 6.0 scenario is a middle of the road scenario assuming technological advances and a leveling off of radiative forcing by 2100. The RCP 8.5 emission scenario is based on high levels of greenhouse gas concentrations and radiative forcing.⁵

These results are summarized in Table 2. The average of the changes in the models imply that the intensity of daily rainfall will increase, which can be seen in the percent increase in the scale parameters from the climate change models. This shifts the rainy day distribution to increase the frequency of larger rainfall amounts. The decreases in the shape parameter implies rainfall on a wet day will be less normally distributed and become closer to an exponentially distributed random variable, in other words: less symmetric and more skewed with more probability mass in the upper tails. These parameter changes mean an increase in the intensity of daily rainfall events. The change in the Markov chain parameters are mixed, but a decrease in the probability of

⁴ We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

⁵ Radiative forcing is the amount of energy that is retained by the atmosphere as a result of greenhouse gases. A great explanation is found in Chandler (2010)

a wet day followed by a wet day, $P(w_t|w_{t-1})$, means that when rainfall occurs it will occur in shorter more intense events over a shorter run of consecutive days than under current rainfall conditions. The probability of a wet day following a dry day, $P(w_t|d_{t-1})$, increases, which means dry spells will be shorter with the exception of August where the probability slightly decreases. In summary, these climate models point to a climate that is wetter in total and with more intense rainfall events than under the current conditions. These more intense events should cause more overland flow, sediment erosion, leading to nutrient loading at the reservoir. We rely on the SWAT model to simulate the time series nature of soil saturation, runoff at rainfall events, as well as nutrients transported and associated water flow to the reservoir.

Results

The simulation and optimization provide the least-cost allocation of BMPs and those induced by the policy options in the watershed as well as the resultant distribution of monthly total P concentrations at the reservoir. We compare the allocations in terms of cost, density of practices and size of participating farms. The following results are broken down into those under the observed rainfall scenario followed by those under the climate change rainfall scenarios.

Baseline Rainfall Scenario

The range of outcomes as a result of BMPs installed on pasture land in the watershed is in Figure 2. The minimum BMP adoption line (thin black line) shows the reservoir's P concentration exceeding the 50 ppb 18% of the summer months (May-October), while at full BMP adoption (thick black line) exceeding this threshold 12% of summer

months. The distance between the two lines in Figure 2 at the 50 ppb threshold represents the whole range of feasible outcomes as a result of pasture land BMP implementation in this watershed.

Figure 3 has this range as the x-axis and plots the cost of achieving these risk levels under the various policy assumptions. The black line represents the results of the optimization routine for the risk-based objective, showing the least-cost allocation to meet the various risk levels, which is the efficient frontier. The light dashed line is the outcome of the installation costs for the location-specific subsidy based on an individual farm's average annual loading. This perfectly corresponds with the efficient allocation when the objective is to control mean P concentrations. The resultant risk levels from those allocations are calculated and plotted together. The distance between the solid and light dashed line is the possible cost savings in defining a risk-based objective as opposed to controlling the risk of cyanobacteria blooms using mean-based policy.

The solid grey line is the cost at various risk levels for the flat-payment scenario where the payments offered is based on the relative effectiveness of each BMP, but with no regard to the heterogeneous nature of the various pasture land's emissions. Again, the distance between each line and the efficient frontier represents the additional cost to meeting various risk levels above the cost of efficient allocations. The black points are the cost of the uniform standards, which requires all farmers to adopt the same BMPs, for each of the eight possible combinations of practices.

The relative cost efficiency of the various policies differs across risk levels, which can be seen by the distance between the various alternative policies and the efficient frontier changing over the range. The allocations resulting from location-specific subsidy, taking advantage of locations' differentiated effects, are the closest to the efficient frontier inducing allocations similar to the least-cost allocations. The remaining difference represents the efficiency of addressing the risk-based objective explicitly.

Take an example of achieving a risk level of 13%. The cost of moving from the current state of 18% to 13% could be achieved optimally at \$300,000, but at \$335,000 if allocations were designed to meet mean concentration goals instead. This is a savings of roughly 12% by adopting allocations based on the risk-based objective. The potential cost savings of explicitly adopting a risk-based allocation ranges from 0 to 25%, with an average of 6.5%.

We have established that the efficient allocation of BMPs across the watershed differs in cost for the risk-based goal versus a mean-based goal. Interestingly, BMP adoption strategies for the two approaches differ. The efficient set of BMPs under the risk-based approach is more targeted, with more intensive controls placed on fewer acres. By comparing the differences in allocations, we can show which farms and practices change at each risk level. Figure 4 shows the differences in total acres in each practice. Noticeably, the optimal allocations have fewer acres in practices in total. In terms of total acres of installed practices, the risk-based objective results in more targeted effort, with fewer acres enrolled. This varies, but the number of acres averages 3% less while increasing to 15% less in the higher effort and lower risk levels. The biggest

differences are in the nutrient management practice, a relatively cheap and less effective practice, with fewer enrolled acres under efficient allocations to meet a risk-based goal.

An additional way to see the differences is using a heat map approach (Wilkinson & Friendly 2009), where we characterize the discrete differences based on farm size between the second best and least-cost allocations. Figure 5 shows the discrete changes between the two allocations, sorted and shaded by their relative size. Each rectangle represents a binary decision for a BMP on a farm. The farms are sorted left to right by size from smallest to largest. The darker color implies a larger farm with a practice installed in the risk-based allocation but not under the mean-based allocation. The lighter color implies a larger farm without that BMP in the risk-based allocation, but existing in the mean-based allocation. The neutral color means the BMPs are the same in both allocations. There is some variation across the risk levels, but in general we see the shift to fewer farms in more expensive and impactful practices and a shift from nutrient management (less impactful, less expensive practice) to manure management (more impactful, more expensive practice). We also see many smaller farms shifted out of the risk-based allocations to BMPs on fewer larger farms. The lower the risk level, the more farms switch their practices, which means the differences are largest in allocations near the upper limit of what is achievable for risk reductions.

Given that P loading from pasture lands only made up 7% of the total modeled loading to the reservoir, the range of reduction in mean P concentrations from installing practices (23.9 to 27.3 ppb) seems small and the costs hard to justify depending on

what damages might be associated with changes in mean concentrations. However, depending on the damage function, the effect of nutrient control efforts may look more promising in having an overall impact on the water quality by lowering the chances a water body supports damaging events. For example, over a 20-year timeframe, the range of effects in terms of risk levels means moving the reservoir from 21.6 summer months to 14.4 months with P concentrations above the 50 ppb threshold.

While we treated risk based on strict threshold violations, it may be the case that the shape of the distribution of conditions below the threshold matters for other water quality goals. It is informative to see the tradeoff of targeting the risk-based objective in terms of other moments of the distribution of conditions. There exists a tradeoff between addressing the risk levels efficiently versus the mean concentrations. The shift towards the addressing the risk levels most cost-effectively comes at the expense of a slightly higher mean concentration at each risk level when the allocations differ. This difference is small and between 0, where the allocations match, and .03 ppb at the maximum difference in spatial allocations of BMPs. So, while there is a difference, shifting policy to address a risk-based goal is nearly costless in terms of increased mean concentrations.

We showed that the type of objective, mean or risk-based, matters to the cost-effective adoption of BMPs across farms and the magnitude of the efficiency of policy designs. The risk-based objective leads to more intensive control on fewer acres, and a shift from more spread out effort to fewer, but larger farms with installed BMPs. Our project was focused on BMPs, where the cost structure had fixed and per acre variable cost components. As a result, the size of the farm to which the policy is aimed matters

to the overall cost, with larger farms having generally a smaller per unit cost of P reduction all else equal. This would also be the case where significant economies of scales occur for pollution control measures. Economy of scales applies to the cost of controlling not simply the mean but also higher order moments of a farm's runoff distribution. This lead to a trend of targeted and concentrated effort on fewer but larger farms. Therefore, policy that is spread out more evenly, standards, for example, fared worse.

Climate Scenario Results

Figure 6 shows the cumulative density function of monthly P concentrations in the reservoir for both the baseline and the RCP 6.0 rainfall scenarios. The black lines represent the effects of the maximum (thick black line) and minimum (thin black line) allocation of BMPs under the RCP 6.0 climate change scenario. The thin lines are the effects of the maximum (thick grey line) and minimum (thin grey line) allocation but under the baseline climate scenario. At the 50 ppb threshold, the minimum BMP allocation risk level is 24% and decreases to just over 19% of months exceeding the threshold. This compares to a range of 18-12% under the baseline rainfall scenario. Under future rainfall scenarios, the P conditions that support algal blooms will be heightened and the risk profile shifted. There is no overlap in risk levels achievable between the baseline and climate scenarios based on the available land for management actions. Returning to near the status quo risk level of 18%, as under the baseline climate scenario, would require near maximum effort (BMP installations) under climate change rainfall scenarios.

Figure 7 shows the cost of achieving risk levels under the climate change scenario using the same second best policies as discussed in the baseline rainfall scenario. The cost ranges are similar between the baseline and climate scenarios, but only higher level of risk are achievable. Maintaining risk levels close to those of the current climate would be just above the upper end of the achievable range and near the highest total cost of \$704 thousand.

There exists larger savings from specifying the risk-based objective explicitly over the mean-based objective in the climate change scenarios. There are larger cost reductions moving to the efficient frontier, varying through the levels, than found under baseline rainfall scenarios. The larger total gains are made in the upper limits of what is achievable in terms of risk reduction, with an average cost savings in percentage terms of 10% and ranging from 0-46% depending on the risk level. This difference is larger in mean and range than the difference under the baseline rainfall scenario. The cost savings over the flat subsidy policy design is large, averaging 40%, under efficient allocations, growing in total magnitude over the risk levels. The standards became even less efficient under the climate scenarios.

Comparing optimal allocations between baseline and the RCP 6.0 climate scenario, the allocations are similar in terms of the progression of practices installed and the relative amount of each through the risk levels. The climate scenarios do not imply a major difference in efficient policy or allocations, but imply larger gains to targeting the risk-based objective explicitly, with larger average gains and a larger maximum gain, 46% versus 25%. As rainfall becomes more intense, the possible benefits from moving towards efficient policy based on a probabilistic objective become larger and uniform

standards become progressively less desirable. Given the differences in allocations, with a shift to fewer but larger farms with BMPs installed when addressing a risk-based objective, the results of the climate change scenario suggests the gains from doing so are larger. Under future climate conditions, spatial prioritization may become more beneficial for this threshold based water quality objective.

We presented a middle of the road emissions pathway, RCP 6.0, above. From Table 2 we can see the changes under a more severe scenario of climate change. The same trends in the cost effectiveness are intensified under the RCP 8.5 scenario. Figure 8 shows the cost at risk levels for the baseline and two climate change scenarios together. The risk levels for the two climate scenarios overlap, so a direct comparison at risk levels can be made. For a given risk level, the allocations under the RCP 8.5 scenario are further up the cost curves and result in more BMP implementation. Equivalent risk levels are achieved at costs \$100-500 thousand more than under RCP 6.0 scenario, visually, the distance between the two curves in Figure 8.

Conclusion

The variability of natural systems poses challenges in addressing and quantifying the costs and benefits of policies. This paper uses a combined hydrologic and economic model to solve the discrete optimization problem to compare the cost effectiveness of policies for nutrient loading to the efficient frontier under both a risk-based and a mean-based water quality objective. Controlling a risk factor for cyanobacteria blooms in drinking water, which are dependent on conditions violating a eutrophic threshold, motivated the approach.

Achieving a risk-based water quality goal, especially applicable to cases where damages are convex to a pollutant or are based on a threshold, results in a different management strategy than meeting objectives based on average conditions. Different sets of pollution control actions and allocations across space can affect higher order moments of the distribution of conditions as opposed to just the mean. We find that achieving the risk-based objective leads to a more targeted approach, with more intensive measures being adopted on fewer but larger sites. Given the results above, focusing on fewer, but more important larger farms may be more cost effective in controlling episodic loadings and damages. This is in contrast to traditional approaches of uniform technology standards or spatially uniform incentives.

Policy designs vary in the amount of differentiation in terms of how they treat heterogeneous sources of emissions. For a risk-based objective, the importance of spatial prioritization becomes larger. In our application, the efficiency of three alternative incentive policies showed the relative cost effectiveness of various second best policies to the risk-based efficient allocations. Utilizing location-specific mean loadings, and a watershed wide uniform payment for P abatement, matched the optimal allocations to address a mean concentration goal. This was closest to the risk-based efficient frontier out of the alternatives considered but still up to 25% more costly to meet a given risk level than the efficient allocations.

Rainfall, which is the transportation mechanism for many pollutants affecting water quality, is projected to change in its intensity and timing with climate change. This will affect both the mean and higher order moments of the distribution of water born pollution conditions in the environment. We find this has implications to the efficiency

of policy. Specifically, the climate change scenarios increase the cost difference between mean and risk-based policies. Incorporating location specific factors into policy when the damages are based on a threshold is more important under more intense future rainfall. For our application, this difference grows with up to 46% lost between the efficient and second best policies. The non-location differentiated policies and uniform standards are also progressively worse in terms of efficiency.

Depending on how often and how far away from the threshold a water body of interest is, marginal changes in effort can have effects on the risk profile of a water body. To best make use of what might be a limited area of land where nutrient control effort might take place, prioritizing locations based on addressing the objective that best represents the damage or goal of the policy could save scarce resources. For the case of cyanobacteria blooms, we find that policy designed to address the probability of crossing a monthly eutrophic threshold of phosphorus concentrations differ than that to address mean concentrations, resulting in more concentrated actions being taken on a smaller number of sites.

Acknowledgments

The authors gratefully acknowledge the funding for this project from U. S. Department of Agriculture through Ecosystem Services Program (Grant # 2010-65615-20669) of the National Institute of Food and Agriculture (NIFA), Agricultural and Food Research Initiative (AFRI). Emi Uchida provided fantastic guidance as well as conference resources. We would also like to acknowledge the SWAT model build and calibration work of Tingting Liu, Q Kellog and Art Gold, as well as Brendan

Hammatt for providing valuable help in compiling the climate model data.

REFERENCES

- Arabi, M. et al., 2008. Representation of agricultural conservation practices with SWAT. *Hydrological Processes*, 22(16), pp.3042–3055.
- Auffhammer, M. et al., 2013. Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *Review of Environmental Economics and Policy*, 7(2), pp.181–198.
- Carlson, R.E., 1977. A trophic state index for lakes. *Limnology and oceanography*, 22(2), pp.361–369.
- Chandler, D., 2010. Explained: Radiative forcing. *MIT News*. Available at: <http://newsoffice.mit.edu/2010/explained-radforce-0309> [Accessed May 11, 2015].
- Chesapeake Bay Commission, 2012. Nutrient Credit Trading for the Chesapeake Bay.
- Elofsson, K., 2003. Cost-effective reductions of stochastic agricultural loads to the Baltic Sea. *Ecological Economics*, 47(1), pp.13–31.
- EPA, 2012. Cyanobacteria and Cyanotoxins: Information for Drinking Water Systems.
- ESS Group, Inc., 2011. Regulating Reservoir Limnological Study and Recommendations.
- Gassman, P.W. et al., 2007. The Soil and Water Assessment Tool: Historical development, applications, and future research directions.
- Geng, S., Penning de Vries, F.W.T. & Supit, I., 1986. A simple method for generating daily rainfall data. *Agricultural and Forest Meteorology*, 36(4), pp.363–376.
- Gitau, M.W., Veith, T.L. & Gburek, W.J., 2004. Farm-level optimization of BMP placement for cost-effective pollution reduction. *Transactions of the ASAE*, 47(6), pp.1923–1931.
- Goetz, R.U. & Zilberman, D., 2000. The dynamics of spatial pollution: The case of phosphorus runoff from agricultural land. *Journal of Economic Dynamics and Control*, 24(1), pp.143–163.
- Gren, I.-M., Destouni, G. & Tempone, R., 2002. Cost effective policies for alternative distributions of stochastic water pollution. *Journal of Environmental Management*, 66(2), pp.145–157.
- Harmel, D. et al., 2006. *Compilation of measured nutrient load data for agricultural land uses in the United States*, Wiley Online Library. Available at:

<http://onlinelibrary.wiley.com/doi/10.1111/j.1752-1688.2006.tb05292.x/abstract> [Accessed June 12, 2015].

- Haygarth, P.M. et al., 2005. The phosphorus transfer continuum: Linking source to impact with an interdisciplinary and multi-scaled approach. *Science of The Total Environment*, 344(1–3), pp.5–14.
- AL Heathwaite, R.D., 2000. Characterizing phosphorus loss in surface and subsurface hydrological pathways. *The Science of the total environment*, 251-252, pp.523–38.
- Huang, Y.L. et al., 2012. Simulation-based inexact chance-constrained nonlinear programming for eutrophication management in the Xiangxi Bay of Three Gorges Reservoir. *Journal of Environmental Management*, 108, pp.54–65.
- Hudnell, K., 2008. *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs*, Available at: <http://www.springer.com/biomed/neuroscience/book/978-0-387-75864-0> [Accessed November 13, 2014].
- IPCC, 2007. *Climate Change 2007: impacts, adaptation and vulnerability: contribution of Working Group II to the fourth assessment report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.
- IPCC, 2011. Special Report on Managing Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX). Available at: <http://ipcc-wg2.gov/SREX/report/> [Accessed January 20, 2014].
- James, P.A. & Alexander, R.W., 1998. Soil erosion and runoff in improved pastures of the Clwydian Range, North Wales. *The Journal of Agricultural Science*, 130(04), pp.473–488.
- Jha, M. et al., 2009. *Optimal placement of conservation practices using genetic algorithm with SWAT*, Center for Agricultural and Rural Development, Iowa State University.
- Kampas, A. & White, B., 2004. Administrative Costs and Instrument Choice for Stochastic Non-point Source Pollutants.
- Karl, T. et al., 2009. *Global Climate Change Impacts in the United States.*, Cambridge University Press.
- Kataria, M., Elofsson, K. & Hasler, B., 2010. Distributional Assumptions in Chance-Constrained Programming Models of Stochastic Water Pollution. *Environmental Modeling & Assessment*, 15(4), pp.273–281.

- Kling, C.L., 2011. Economic incentives to improve water quality in agricultural landscapes: Some new variations on old ideas. *American Journal of Agricultural Economics*, 93(2), pp.297–309.
- Kunkel, K.E. et al., 2013. Probable maximum precipitation and climate change: PMP AND CLIMATE CHANGE. *Geophysical Research Letters*, 40(7), pp.1402–1408.
- Lathrop, R.C. et al., 1998. Phosphorus loading reductions needed to control blue-green algal blooms in Lake Mendota. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(5), pp.1169–1178.
- Meehl, G.A., Arblaster, J.M. & Tebaldi, C., 2005. Understanding future patterns of increased precipitation intensity in climate model simulations: Patterns of precipitation extremes. *Geophysical Research Letters*, 32(18).
- Milon, J.W., 1987. *Optimizing Nonpoint Source Controls in Water Quality Regulation*, Wiley Online Library.
- Morgan, R.P.C., 2001. A simple approach to soil loss prediction: a revised Morgan–Morgan–Finney model. *CATENA*, 44(4), pp.305–322.
- Qui, 2011. SWAT Modeling Analysis for the Neshanic River Watershed.
- Rabotyagov, S. et al., 2010. Least-cost control of agricultural nutrient contributions to the Gulf of Mexico hypoxic zone. *Ecological Applications*, 20(6), pp.1542–1555.
- Rao, N.S. et al., 2009. Modeling watershed-scale effectiveness of agricultural best management practices to reduce phosphorus loading. *Journal of Environmental Management*, 90(3), pp.1385–1395.
- RCP Database, 2014. Available at:
<http://tntcat.iiasa.ac.at/RcpDb/dsd?Action=htmlpage&page=welcome#intro>
 [Accessed November 18, 2014].
- Reclamation, 2013. Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections. Available at: http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/dcpInterface.html [Accessed November 13, 2014].
- Rhode Island Office of Water Resources, 2009. Water Quality Regulations.
- Segerson, K., 1988. Uncertainty and Incentives for Nonpoint Pollution Control. *Journal of Environmental Economics and Management*, 15, pp.87–98.
- Shortle, J., 1990. The Allocative Efficiency Implications of Water Pollution Abatement Cost Comparisons. *Water Resources Research*, 26(5), pp.793–797.

- Shortle, J. & Horan, R.D., 2013. Policy Instruments for Water Quality Protection. *Annual Review of Resource Economics*, 5(1), pp.111–138.
- Shortle, J.S. & Horan, R.D., 2001. The economics of nonpoint pollution control. *Journal of Economic Surveys*, 15(3), pp.255–289.
- Srikanthan, R. & McMahon, T.A., 1999. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, 5(4), pp.653–670.
- Srivastava, P. et al., 2002. Watershed optimization of best management practices using AnnAGNPS and a genetic algorithm: watershed optimization. *Water Resources Research*, 38(3), pp.3–1–3–14.
- USDA, 2015. Conservation Programs. Available at: <http://www.fsa.usda.gov/programs-and-services/conservation-programs/index> [Accessed June 10, 2015].
- USDA Natural Resource Conservation Service, 2012. RI FY2012 Practice Payment Schedules. Available at: <http://efotg.sc.egov.usda.gov/treemenuFS.aspx>.
- van Dijk, A.I.J.M., Bruijnzeel, L.A. & Rosewell, C.J., 2002. Rainfall intensity–kinetic energy relationships: a critical literature appraisal. *Journal of Hydrology*, 261(1–4), pp.1–23.
- Vollenweider, R.A., 1971. *Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication*, Organization for Economic Co-operation and Development Paris.
- Wilkinson, L. & Friendly, M., 2009. The history of the cluster heat map. *The American Statistician*, 63(2). Available at: <http://amstat.tandfonline.com/doi/abs/10.1198/tas.2009.0033> [Accessed June 11, 2015].
- Wilks, D.S., 2012. Stochastic weather generators for climate-change downscaling, part II: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3), pp.267–278.
- Zhu, M. et al., 1994. Chance constrained programming models for risk-based economic and policy analysis of soil conservation. *Agricultural and Resource Economics Review*, 23(1), pp.58–65.

Figures and Tables

TABLE 1
BMP Characteristics

BMP	P reduction efficiency	Fixed Cost	Variable Cost	
Prescribed Grazing	25%	\$3,750	\$266	per acre
Manure Management	80%	\$1,000	\$2,300	per acre
Nutrient Management	47%	\$1,650	\$75	per acre

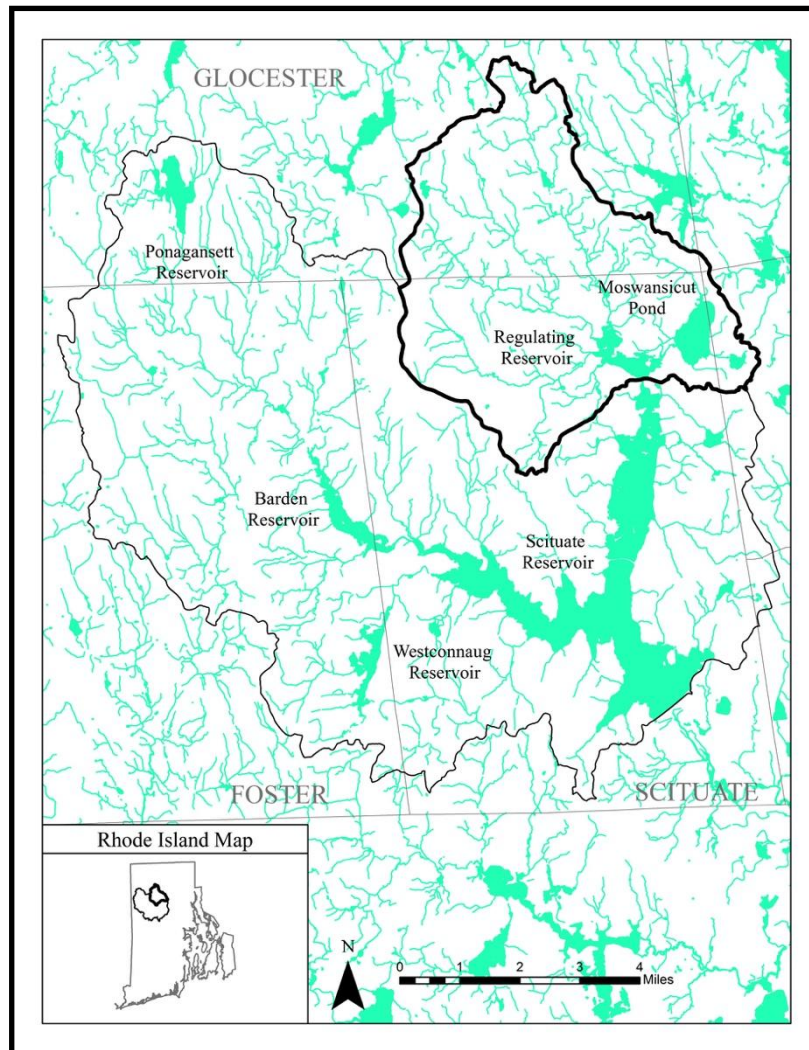
NOTE- BMP efficiencies were adapted from Chesapeake Bay Watershed Model (Chesapeake Bay Commission 2012; Gitau et al. 2004; Qui 2011). Cost information was adapted from NRCS Equip costing information (USDA Natural Resource Conservation Service 2012). These practices are the ones deemed relevant to the RI study area. Manure management practice was scaled to reflect the watershed animal density and acreage to create the variable cost.

TABLE 2
Climate Change Rainfall Parameters

Historic Rainfall Parameters					Climate Adjustment (RCP 6.0)				Climate Adjustment (RCP 8.5)			
Gamma Parameters			Markov Probabilities		Gamma Parameters		Markov Probabilities		Gamma Parameters		Markov Probabilities	
Month	Scale (mm)	Shape (mm)	P(Wt Wt-1)	P(Wt Dt-1)	Scale	Shape	P(Wt Wt-1)	P(Wt Dt-1)	Scale	Shape	P(Wt Wt-1)	P(Wt Dt-1)
Jan	13.00	0.46	0.40	0.16	19.2%	-8.8%	-1.9%	5.8%	25.0%	-9.2%	-1.0%	2.7%
Feb	13.01	0.46	0.48	0.16	12.6%	1.5%	0.4%	-0.9%	19.3%	-3.0%	0.4%	0.1%
Mar	15.57	0.41	0.45	0.15	14.0%	-1.2%	0.8%	-1.7%	20.3%	-2.3%	-0.2%	1.0%
Apr	15.85	0.38	0.45	0.15	10.5%	1.8%	-0.1%	0.5%	19.6%	-3.2%	-2.7%	7.3%
May	12.26	0.42	0.44	0.15	12.6%	-5.1%	-3.3%	11.6%	15.7%	-5.4%	-3.6%	11.3%
Jun	21.85	0.21	0.47	0.14	13.3%	-2.4%	-2.4%	7.0%	17.2%	-4.3%	-3.1%	10.3%
Jul	15.73	0.34	0.38	0.15	9.4%	6.4%	-2.8%	8.3%	14.7%	-0.1%	0.0%	-0.8%
Aug	23.61	0.27	0.42	0.15	13.2%	4.2%	0.4%	-1.3%	11.6%	5.9%	0.8%	-4.1%
Sep	21.09	0.33	0.35	0.15	2.7%	6.7%	-3.9%	5.7%	6.7%	5.2%	-2.9%	2.5%
Oct	18.13	0.39	0.32	0.13	14.7%	-7.5%	-9.2%	10.1%	13.7%	-9.9%	-8.4%	10.3%
Nov	16.93	0.43	0.41	0.15	13.6%	-1.5%	-3.7%	9.2%	19.3%	-7.7%	-5.3%	10.5%
Dec	15.65	0.40	0.45	0.16	17.3%	-4.0%	-1.8%	4.1%	23.5%	-7.0%	-2.4%	6.3%

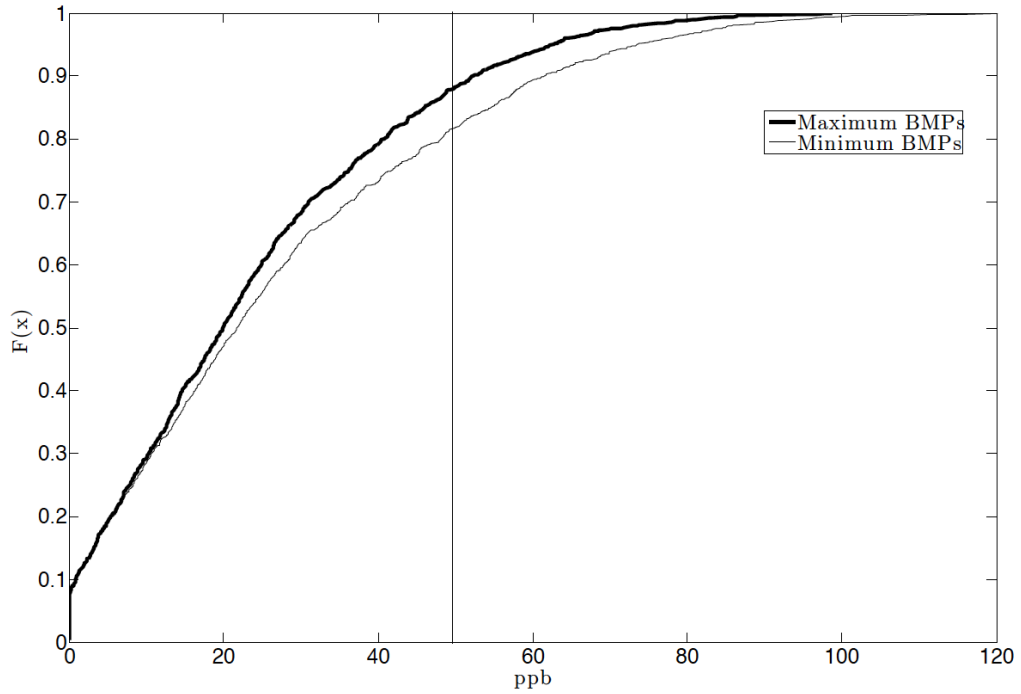
NOTE- Gamma parameters, shape and scale, as well as the Markov chain parameters of wet days were fit to observations from for 1950-1999 by month in mm of rainfall from the Kingston RI weather station. The differences between past (1950-1999) and future (2050-2099) runs of climate models (both RCP 6.0 and RCP 8.5 emission paths) for these same parameters are presented as adjustments. These % changes were applied to the observed parameters that were then used to create the time series of future rainfall for climate scenarios. The climate model runs were from the Bureau of Reclamation's Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections archive (Reclamation, 2013)

FIGURE 1
Scituate Reservoir Watershed, RI



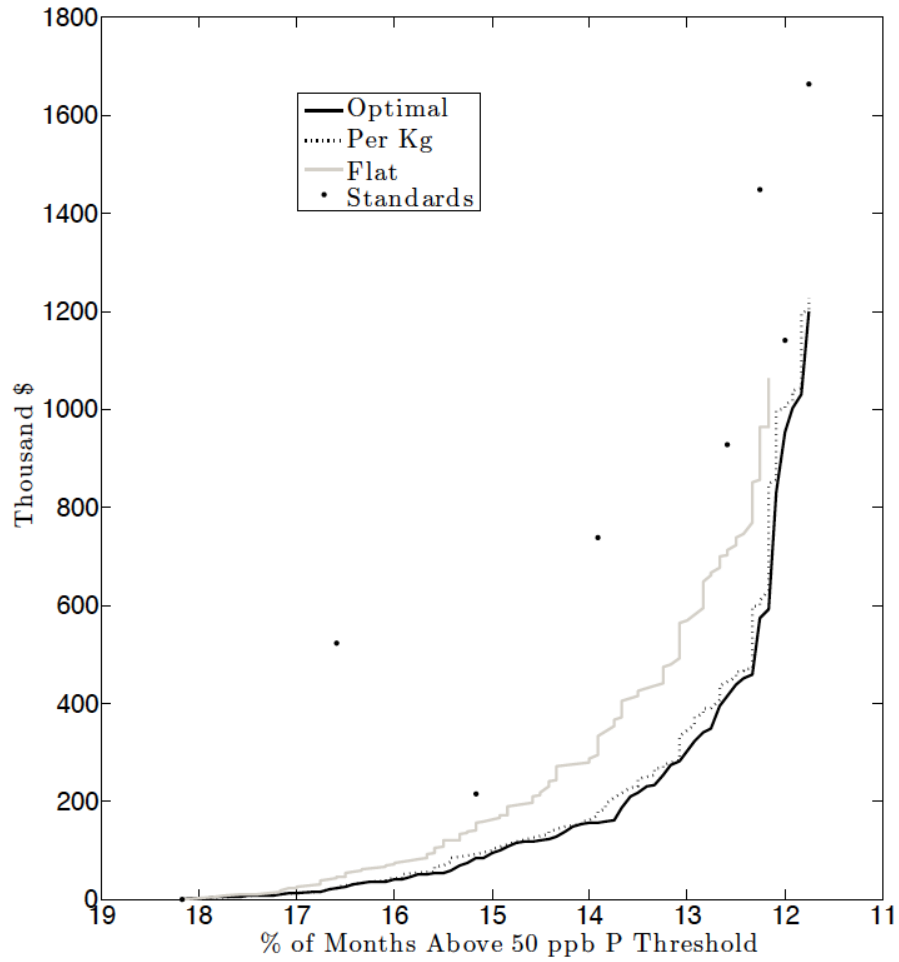
NOTE – The Scituate Reservoir Watershed is 86 square miles and is made up of five sub-watersheds, which drain into the Scituate Reservoir system, which is owned and managed by the Providence Water Supply Board (PWSB). The reservoir system provides water services to 60% of residents and business of Rhode Island including the municipalities of Providence, N. Providence, Warwick, E. Providence, Western Cranston, Kent Co., East Smithfield, Smithfield, Greenville and Lincoln accounting for 92 million gallons a day (Scituate Reservoir Source Water Assessment 2003). The Regulating reservoir watershed is highlighted in black.

FIGURE 2
CDF of Monthly P Concentrations at the Reservoir
Baseline Rainfall Scenario



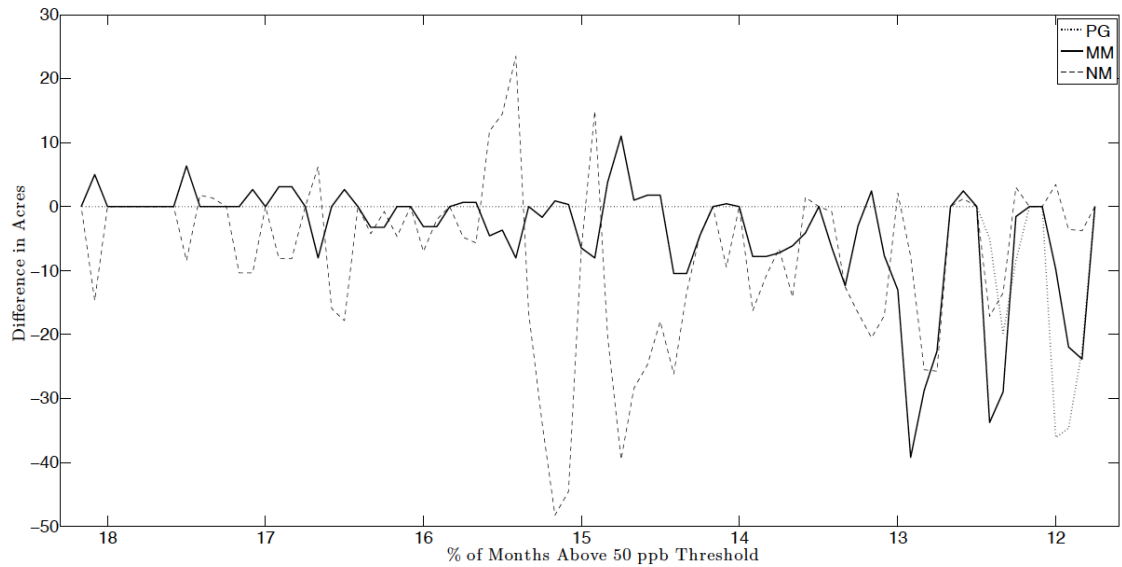
NOTE- The x-axis represents monthly P concentrations of inflows to the reservoir. The y-axis is the probability that the reservoir is below that concentration. The two lines represent the outcome of the minimum and maximum BMP implementation on pasture land in the watershed. The vertical line is the 50 ppb threshold used as an indicator of a high risk of algae blooms. The distance between the two lines at the threshold is the feasible risk levels, which is the x-axis in Figure 3.

FIGURE 3
Total Costs at Risk Levels
Baseline Rainfall Scenario



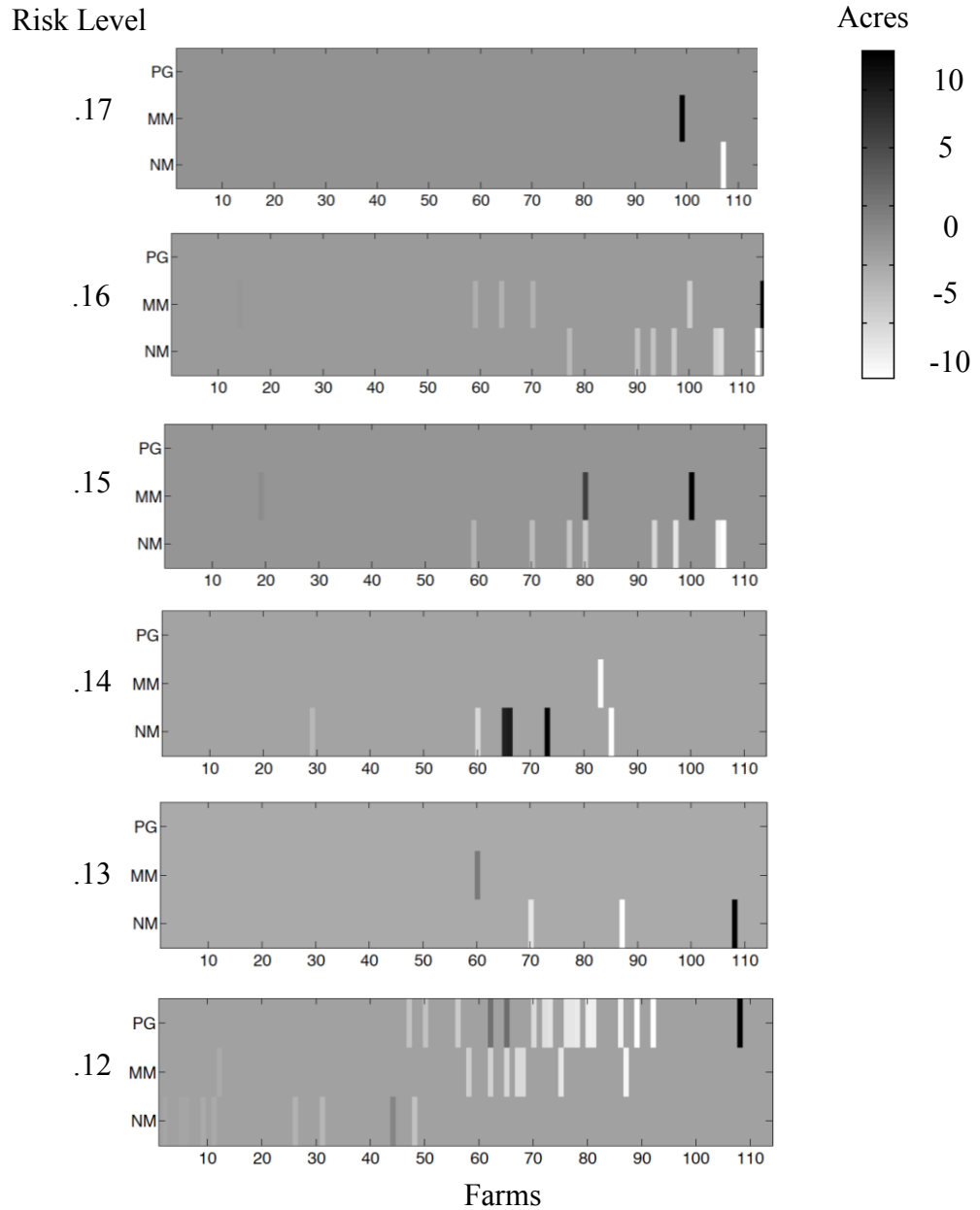
NOTE- The solid lines represent the cost of achieving risk levels assuming optimal policy (solid black), a payment based on the yearly average P reductions as a result of installed practices (dashed), a flat per acre payment based on the relative BMP efficiencies (solid grey), but not taking into account other site specific characteristics. The black dots represent the risk levels and costs assuming standards.

FIGURE 4
Differences in Acres with BMPs
Risk-Based Compared to Mean-Based



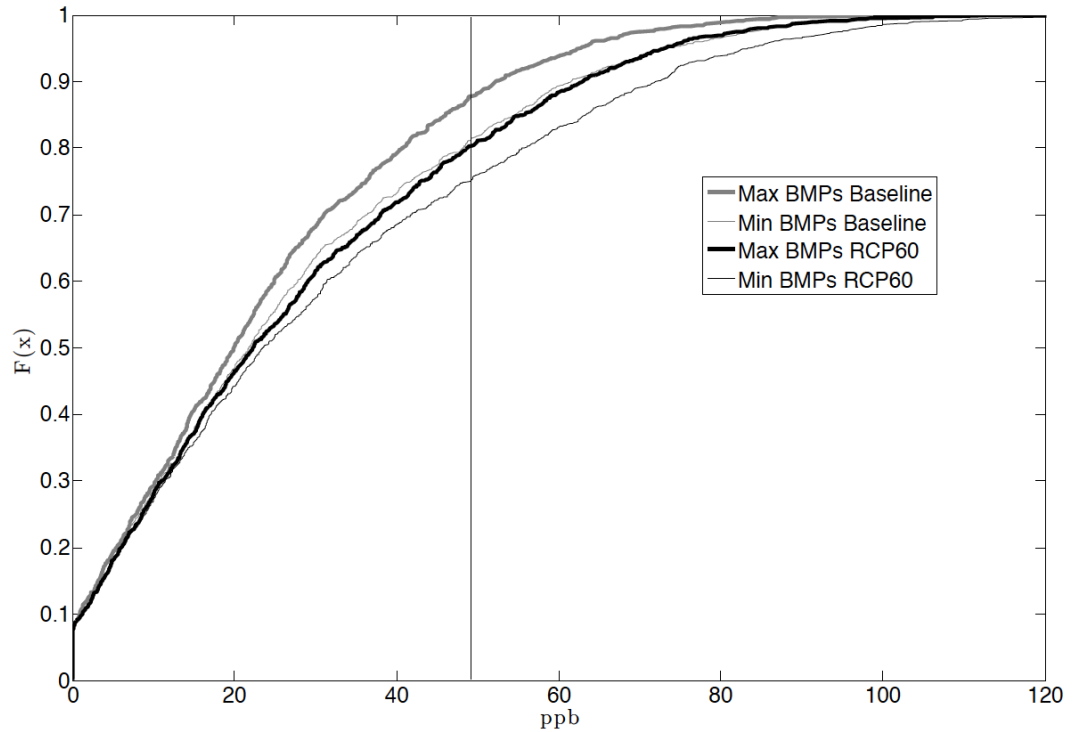
NOTE- The differences in the total amount of acres in the three practices (prescribed grazing, manure management, nutrient management) are shown between the monthly risk and mean-based objective. The positive acres represent additional acres in practices for optimally addressing the risk-based objective. The zero line shows the two allocations of BMPs to be the same and negative implies more acres in practices than under the mean loading objective.

FIGURE 5
Discrete Differences in Installed BMPs
Risk Versus Mean Allocations



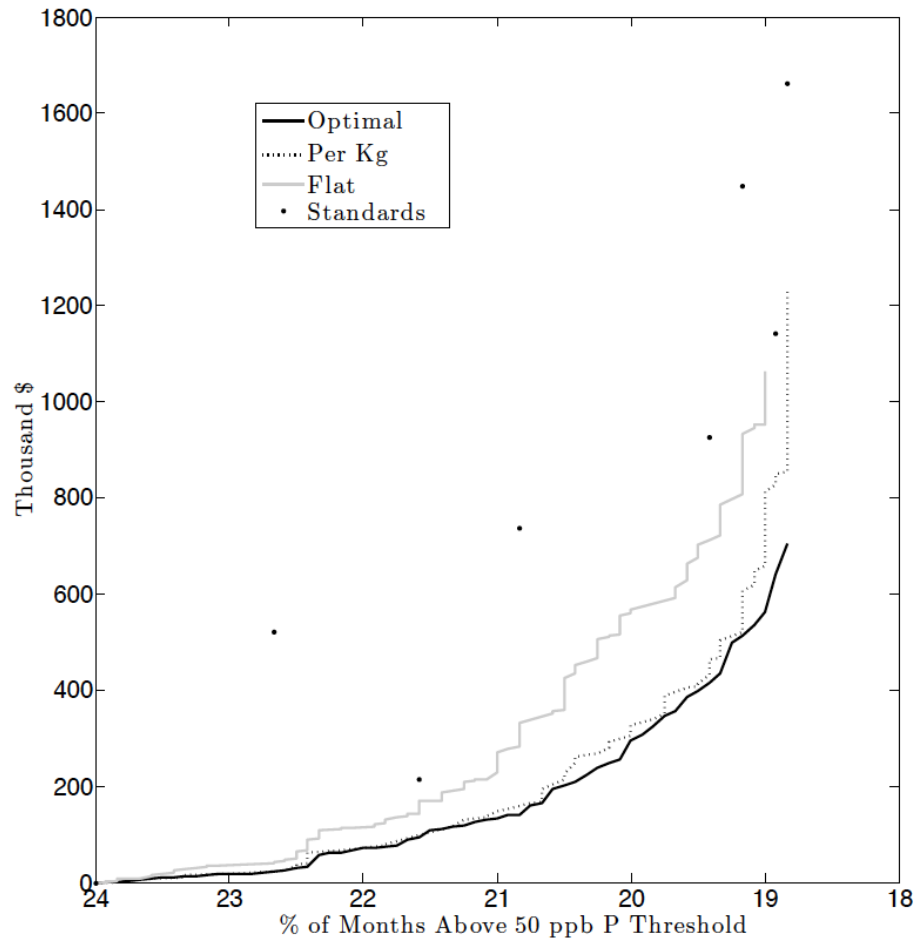
NOTE- Discrete changes between the two allocations at various risk levels, sorted and shaded by their relative size. Each block represents a discrete decision on a farm. Darker color implies a larger farm added to the risk-based allocation and whiter color shows a larger farm subtracted from the risk-based objective, or alternatively thought of as existing in that practice under the mean-based allocation, but not under the risk-based allocations. There is a shift towards more concentrated BMPs on larger farms under the monthly risk-based objective.

FIGURE 6
CDF of Monthly P Concentrations at Reservoir
RCP 6.0 Rainfall Scenario



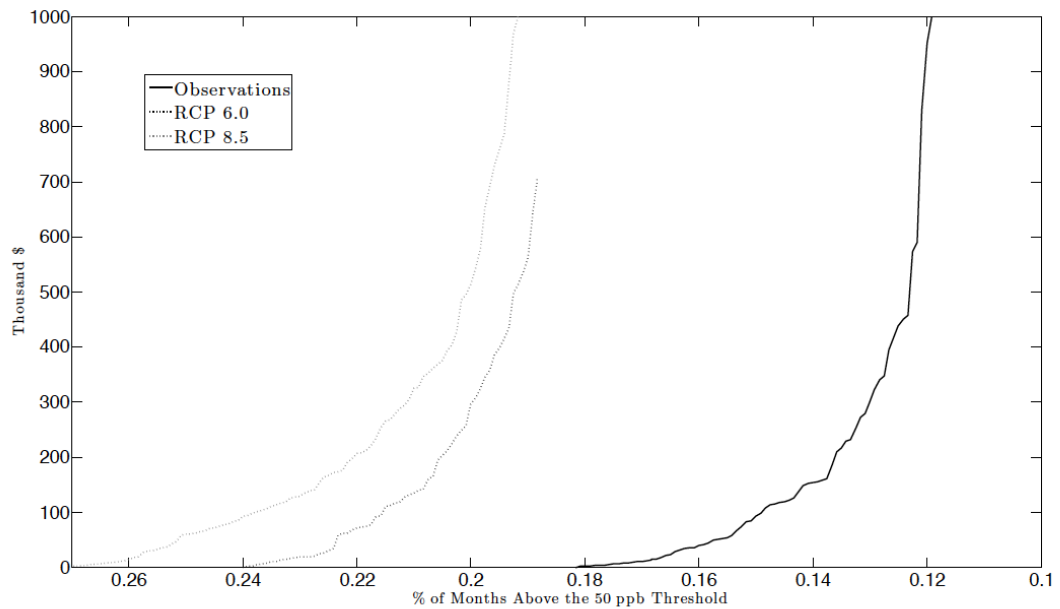
NOTE- The x-axis represents monthly P concentrations of inflows to the reservoir. The two pairs of lines represent the minimum (thin) and maximum (thick) effects of BMP implementation on pasture land in the watershed. The grey lines are from Figure 2 and represent the baseline climate scenario, while the black lines are from runs of the RCP 6.0 climate change scenario. The climate scenarios show increased loadings and concentrations above the threshold. The vertical line is the 50 ppb threshold used as an indicator of a high risk of algae blooms. The distance between the two lines at the threshold is the feasible risk levels, which is the x-axis in Figure 7.

FIGURE 7
Total Costs at Risk Levels
RCP 6.0 Rainfall Scenario



NOTE- The black solid lines represent the cost of achieving risk levels assuming optimal policy (black), a payment based on the yearly average P reductions as a result of installed practices (dashed line), a flat per acre payment based on the relative BMP efficiencies (grey line), but not taking into account other site specific characteristics. The black dots represent the risk levels and costs assuming blanket standards.

FIGURE 8
Total Cost Curves Under Baseline and Climate Change Rainfall Scenarios
Optimal BMP Allocations



NOTE- This figure shows the minimum cost of achieving risk levels for the baseline and the two climate change scenarios of rainfall parameters. The black line represents the cost curve under observed rainfall parameters from 1961-2010 from the Kingston, RI weather station. The dashed lines represent the least-cost curves resulting from future projections of rainfall for the years 2050-2099 for the RCP 6.0 and RCP 8.5 scenarios.

Manuscript 2

Optimal Groundwater Extraction Under Uncertainty and a Spatial Stock Externality

By

Nathaniel H. Merrill, Todd Guilfoos

Department of Environmental and Natural Resource Economics

University of Rhode Island

Submitted for Publication to the *Journal of Environmental Economics and Management* (Under Review)

Optimal Groundwater Extraction Under Uncertainty and a Spatial Stock Externality

Nathaniel H. Merrill^a, Todd Guilfoos^b

^a University of Rhode Island, Email: nateme16@my.uri.edu

^b University of Rhode Island, Email: guilfoos@mail.uri.edu

Department of Environmental and Natural Resource Economics
University of Rhode Island
Coastal Institute
1 Greenhouse Rd, Kingston RI 02881

Abstract

We introduce a model that incorporates two important elements to groundwater management: stochasticity and a spatial stock externality. We estimate welfare gains resulting from optimal management under uncertainty as well as a gradual stock externality that produces the dynamics of a large aquifer being slowly exhausted. Using dynamic programming, we incorporate and compare stochasticity for both an independent and identically distributed (i.i.d.) as well as a Markov chain process for annual precipitation. We investigate optimal responses for management of groundwater resources and the importance of these responses as measured by welfare gains. We find that the spatial depletion of the aquifer is significant to welfare gains but the inclusion of various forms of uncertainty in rainfall increases estimates only slightly.

Introduction

Groundwater plays an important role in mitigating the effect of weather variability on economic activity. Agriculture, in particular, is highly dependent on rainfall and is thus uniquely exposed to weather-related risk. Many groundwater aquifers around the world are stable supplies of water for irrigation to make up for deficits in natural rainfall. As more than an additional source of water, they act as a water buffer in low rainfall years. According to the World Bank, it is estimated that 2 billion people worldwide depend on groundwater for drinking water. In addition, 50% of cereal production depends on groundwater for irrigation (Wijnen et al. 2012). As a relatively low cost source of clean water, aquifers are being depleted at an alarming rate in many places around the world. Groundwater depletion has led to longer pumping heights, saline intrusion in coastal systems, and land subsidence (United Nations Environment Programme & Division of Early Warning and Assessment 2003). Aquifers such as the Indus River plains aquifer on the India-Pakistan boarder, the North China Plain aquifer, and the High Plains/Ogallala aquifer in the United States are experiencing declines as society's reliance on groundwater continues to grow (United Nations Environment Programme 2012).

Irrigated agricultural plays an important role in reducing the lifespan of many of these aquifers. For instance, over the Ogallala aquifer, irrigation began in the late 1800's but intensified greatly after WWII with the introduction of pivot irrigation systems. As of 2011, total storage across the aquifer has fallen 8.3% since 1950. This depletion of the resource has not been spatially uniform across the whole system with some limited

areas gaining while most areas have faced falling levels. In some places, decreases as much as 242 feet have been observed, as can be seen in Figure 1 (McGuire 2012). In western Kansas, areas are left with little or no saturated thickness, and farming is forced to transition from irrigated crops, mostly corn, to non-irrigated crops such as sorghum, wheat and cotton (Steward et al. 2013; The Economist 2013). This switch from high value, stable, irrigated crops to lower value dryland varieties is captured by the spatial stock dynamics of our model.

As groundwater is depleted, there is a loss in its marginal value due to higher costs of extraction. This is in addition to a loss of the groundwater stock's marginal impact in reducing the variability of returns to farming (Tsur & Tomasi 1991; Knapp & Olson 1995). Despite this buffer value being a potentially large percentage of the total value of a groundwater resource⁶, welfare gains estimated when moving from open access to optimal behavior are relatively small in size, supporting the findings from Gisser and Sanchez (1980).⁷ A review of research addressing the curious finding of Gisser & Sanchez (1980) by Koundouri (2004) presents the range of assumptions and the welfare gains from management that have been estimated showing the robustness of Gisser & Sanchez's result across many settings.

Groundwater is a valuable source in times of drought. Harou et al. (2010) use a synthetic drought time series based on historic droughts to quantify the statewide impacts in California based on constraints on optimal monthly flows in a linear

⁶ Buffer values were estimated at between 5-84% of total value, but a comparison between myopic and optimal management was not presented (Tsur and Tomasi 1991)

⁷ Knapp and Olson (1995) find gains to optimal management of 2.6% for Kern County, CA

programming model. They find that the scarcity value of water increases tenfold under drought conditions even when optimally managed, although they do not present results in net present value form. The method implies that the exact time series of rainfall is perfectly known over the optimal planning horizon, which is a stronger assumption than knowledge of the rainfall process, but not the exact realization. Roseta-Palma & Xepapadeas (2004) present a more general formulation and use robust control techniques to estimate optimal groundwater resource use. Precautionary behavior implies that optimal management under uncertainty results in extraction rates that are lower to save for future shortcomings than under more certainty in the random process of rainfall. Welfare effects of optimal versus open access management are not discussed in this framework, and the results are theoretical and not parameterized to a groundwater source. Zeitouni (2004) presents results for an aquifer with stochastic recharge but constant, not rain dependent, demand and constant marginal benefit. This leads to a bang-bang solution for optimal control and a target groundwater level. In contrast, our paper will use constant recharge, crop yield functions and dual sources of water, rainfall and groundwater, in describing the marginal benefit curves, demand, extraction rules, and resulting stock dynamics.

In addition to uncertainty in rainfall and surface water sources, other papers have addressed irreversibility and catastrophic loss of a resource. Tsur & Zemel (2004) and Leizarowitz & Tsur (2012) model the threat of a discrete permanent change in the system both with certain and uncertain stock dynamics. In specifications of stock-dependent event risk, prudence or lower extractions are optimal to save for future

states, although this does not hold for irreversible exogenous events, which imply larger extractions than otherwise. In contrast, our paper does not address discrete event risk or regime changes. We will compare optimal policy and resource savings under specifications in the rainfall process as well as quantifying the implications to the externalities associated with open access usage.

Steward et al. (2013) estimates the effect of scenarios of irrigation pumping reductions on the lifespan of the Ogallala Aquifer as well as crop yields and irrigation supported cattle over the next century. The study finds certain wells are located in areas of thinner saturated thickness which supports the current trend of steady increases in the total yield from dryland compared to irrigated farming across the aquifer. An empirical discussion of the relative returns and drought adaptation of irrigated and dryland farming over the Ogallala can be found in Hornbeck and Keskin (2012). While they argue that crop adaptation in the long run (extensive margin) can mitigate some of the negative consequences of losing access to the aquifer through increased dryland farming, they do not attempt to quantify the welfare impacts when the size of irrigable land is endogenous and returns are calculated aquifer wide.

The spatial and connected nature of aquifers has been treated in a variety of ways in the literature. A review of groundwater representations in the economics literature can be found in Katic (2010). Studies that account for lateral flows of groundwater (Saak & Peterson 2007; Saak & Peterson 2012; Brozović et al. 2010; Savage & Brozović 2011; Palazzo & Brozović 2014; Guilfoos et al. 2013) have explored the under, or

over, representation of the pumping height externality when assuming a uniform single cell aquifer, known as the bathtub model. The specification of this particular externality is meaningful for welfare analysis since it is the magnitude of the pumping height externality, when taken into account in optimal management, which results in welfare gains of optimal extraction. Depending on the spatial and size distribution of users, welfare gains to management can be over or under estimated when using a bathtub type model of groundwater. In addition, the degree of the associated externalities dissipated with increasing degrees of strategic behavior and ownership by extractors, as pointed out by Brozović et al. (2010) and Guilfoos et al. (2013). A different, and possibly larger stock externality, that of the depletion of the total spatial extent of an aquifer, is explored in this paper with a more detailed multi-user spatial approach put aside for clarity. Simply put, in the model presented, as a groundwater resource is depleted, so too is the irrigable land lying above with access to the aquifer. Thus, farm operations on the land that have lost access to groundwater must transition from irrigated crops to dryland crops.⁸

We introduce a new economic groundwater model that incorporates the gradual shift from irrigation to dryland farming as parts of the aquifer run dry. We accomplish this using an upside down cone to represent the spatial depletion, where the area of irrigable land above the aquifer shrinks as the water level decreases. Depletion of the aquifer may interact with uncertainty of the supply of water because the buffer that

⁸ Koundouri & Christou (2006) show that the value of a backstop, in our case dryland farming, has a large effect on the magnitude of the welfare gains from management with the addition of a more valuable backstop attenuating the gains available.

groundwater provided is no longer available. In this work, we identify the impact of spatial depletion on welfare gains from optimal management when rainfall is stochastic and of a Markov process. With this model, we estimate gains from moving away from current myopic extraction behavior to optimal use of the resource. When applied to Kansas over a section of the Ogallala Aquifer, we find gains from management ranging from 2.88% to 3.01% with larger gains achieved under uncertainty in the rainfall process. These results quantify the importance of the spatial externality as well as including uncertainty. In the process, we make two contributions to the literature. First, the introduction of a model that allows for a gradual spatial externality with a backstop technology of dryland farming, in contrast to other works that model an abrupt loss of the entire aquifer (Koundouri & Christou 2006). Second, we estimate the welfare gains from management under various forms of uncertainty to understand the impact of modeling serial correlation in rainfall. We find that including the gradual depletion of the aquifer does materially impact welfare results compared to other estimates of the same region, and surprisingly the serial correlation of rainfall matters little. Empirically, multi-year droughts combined with the loss of access to the aquifer only slightly increases welfare gains due to the availability of dryland farming and the productivity of that option as a backstop when available. The model is presented in detail below, followed by an empirical application and discussion.

Conceptual Model

Large groundwater resources, such as the Ogallala, are examples of common pool resources managed subject to various forms of open access, where, given the institutions in place, any one user does not have the incentive, or information, to internalize the externality of their own pumping in reducing water levels over time as discussed in the literature (Hornbeck & Keskin 2012; Steward et al. 2013; Guilfoos et al. 2013; Saak & Peterson 2007). As such, groundwater is an acute example of resource exploitation leading to possibly larger reliance on variable natural rainfall. In times of drought, this can have large economic consequences. Our model captures the important interaction of lost access to groundwater and exposure to uncertain rainfall events.

When dynamics of the loss of irrigable land are included, the buffer value of groundwater in reducing the variability of profits may be particularly important, as more farmland is transitioned to dryland, rain dependent farming. The extent to which groundwater can act as a buffer against variable future rainfall is also a function of the costs of extraction in a given period, which in this model, is dependent on the stock, or height of the groundwater. Therefore, with higher levels of groundwater, more can be extracted across a larger area of farmland to meet demand in realizations of low rainfall, thus limiting the variability of aquifer-wide profits in any given year. As groundwater levels fall, so too does the amount available for withdrawal in response to low rainfall. Therefore, groundwater depletion increases the variance of possible realizations of aquifer wide profits in subsequent periods. The extent of the welfare

gain due to optimal management is an empirical question, which we address with an application.

Spatial Externality

We model the dynamics of how groundwater evolves and access to the resource changes by introducing a spatial depletion function where the amount of irrigable farmland above the aquifer is a function of the pumping height (Figure 2). This specification represents an upside-down cone shape where the area above the remaining groundwater faces uniform pumping heights, but the area no longer above the groundwater is left without any option for irrigation. Doing so creates an analytical relationship between the amount of water extracted from the aquifer and the evolution of the height of groundwater and the changing access to groundwater. This captures the important heterogeneity in groundwater resources' usable lifespan over the aquifer. Meaningful lifespans for irrigation pumping are a function of various location specific factors, including pumping height and saturated thickness, but also are affected by hydrologic properties of the soils. As such, the chosen function captures the most at risk areas being exhausted sooner. We have simplified the spatial externality of lost irrigable acreage by making the percent of remaining irrigable land a function of simply the pumping height. Equation (1) defines the percentage of irrigable land as a function of height, which is bounded by 0 and 1.

$$\gamma_t(x_t) = \frac{\pi \left(\frac{x_t}{x_0/u} \right)^2}{A} \quad \gamma_t \in [0,1] \quad (1)$$

Where x_0 is the initial groundwater height, u the radius of the circle that is the base of a cone that represents the irrigable land area above the aquifer. Derivation of equation (1) can be found in the Appendix. This specification captures both the heterogeneity in saturated thickness as well as allowing for changes at the extensive margin (switch to dryland practices). This model is most applicable to areas where the marginal cost of increased pumping heights is relatively small compared the value of groundwater, or when the spatial distribution of wells and water demand is relatively uniform. By simplifying this spatial externality, it allows us to investigate the stochastic dynamics while keeping the size of the stochastic dynamic programming problem computationally manageable.

Rainfall Process

There are many ways to model the variability of rainfall, depending on the time and spatial scale appropriate to the problem (Srikanthan & McMahon 1999). We choose to model annual rainfall expectations and realizations two ways: the first scenario assumes independent and identically distributed (i.i.d.) realizations of rainfall. There are n discrete rainfall states, r^n , and corresponding probabilities, p^n . The probability of next year's rainfall, r_{t+1} , is independent of the current rainfall state, r_t , thus $prob(r_{t+1} = r^n) = p^n$.

$$E(r_{t+1}) = \sum_1^n p^n r^n \quad (2)$$

Equation (2) defines expectations when rainfall is i.i.d., where E is the expectation operator.

However, it is clear that rainfall time series are not serially independent and there exists many stochastic rainfall generation methods to model the serial dependence (Srikanthan & McMahon 1999). Therefore, the second scenario presented in this paper assumes a simple Markov chain process in order to replicate climate persistence, particularly droughts, where the probabilities of future rainfall states are a function of the current year's rainfall. This process defines probabilities of the future state that are conditional on the current state, $prob(r_{t+1} = r^n | r_t) = p^n(r_t)$, and

$$E(r_{t+1} | r_t) = \sum_1^n p^n(r_t) r^n \quad (3)$$

By specifying the transition probabilities of moving from one state to another, we capture the time dependent process and the persistence of annual rainfall.

Economic Benefits

In order to estimate the per-acre return to farming, we assumed a section that is irrigated acreage and a section of farmland that is dryland acreage. The returns to irrigated land are a function of per-acre yield, price, irrigation pumping cost and the quantity of irrigation water extracted and applied. Equation (4) gives the per-acre returns to irrigated farmland.

$$f_I(w_t, x_t | r_t) = y_I(w_t, r_t) p_c - c_1 (x_0 - x_t) w_t \quad (4)$$

Where y_I is the per-acre yield, w_t is the irrigation water applied per-acre, p_c is the price of the irrigated crop and c_1 is the lift, $(x_0 - x_t)$, dependent marginal cost of one acre foot of groundwater extraction. x_t is the elevation of the groundwater level as compared the surface elevation, x_0 . Taken together, the cost function gives the marginal cost of pumping one acre-foot of water to the surface for a given pumping height. Myopic farmers that irrigate do so to maximize profits in each time period by choosing w_t to maximize (4). Farmland that does not have access to the aquifer only has rainfall as a water input into the production of crops and is represented by,

$$f_D(r_t) = y_d(r_t) p_d \quad (5)$$

Where $y_d(r_t)$ is the per-acre yield of dryland crops as a function of rainfall and p_d the prices of those crops.

As mentioned before, irrigable farmland is itself a function of pumping height. So a single period's aquifer wide return would be the area-weighted sum of irrigated and dryland profits:

$$\Pi_t = A[\gamma(x_t) f_I(w_t, r_t, x_t | r_t) + (1 - \gamma(x_t)) f_D(r_t)] \quad (6)$$

We assume the homogenous per-acre irrigation water demand across the irrigated portion of the aquifer, so the total water extracted is:

$$W_t = [A\gamma_t(x_t)] w_t \quad (7)$$

Based on the cumulative extraction W_t , the groundwater height changes through time following the equation of motion:

$$g(x, w) = x_{t+1} = x_t + \frac{R - (1 - \alpha)W_t}{\gamma_t(x_t)AS} \quad (8)$$

R is the natural recharge, α the percent of irrigation water returning to the aquifer, A the total area of the aquifer, S the storativity and $\gamma_t(x_t)$ the percent of irrigated land over the initial aquifer area from equation (1). This formulation of the equation of motion has consequences to the rate of depletion. Given a similar per-acre pumping demand, w_t , height changes will be smaller under this spatial depletion model than under the assumption of a traditional bathtub model. This is due to two factors: the loss of irrigated acreage as the heights fall, and the effect of the change in the dimensions of the area that the volume of recharge fills. At lower groundwater levels, a similar volume of recharge will result in a larger positive height change than at higher levels, due to the shape of the function. A more formal discussion of the equation of motion is in the Appendix.

The social planner's problem is to find the extraction path to maximize the discounted sum of future profits, subject to the equation of motion (8) and physical constraints, as given here:

$$\max_{w_t} E_{r_t} \left[\sum_{t=0}^{\infty} \beta^t \Pi_t(w_t, x_t, \gamma_t | r_t) \right] \quad (9)$$

s.t.

$$x_{t+1} = x_t + \frac{R - (1 - \alpha)W_t}{\gamma_t AS}$$

$$\gamma_t = a(x_t)$$

$$\gamma_0 = 1$$

$$x_0 = \bar{x}$$

$$x_t \in [\bar{x}, \underline{x})$$

Where E_{r_t} , is the expected value operator over the stochastic rainfall variable, r_t . β is the discount factor, t the year index and \bar{x} and \underline{x} are the maximum and minimum groundwater levels. The random nature of rainfall is important in this setting as groundwater is used to augment natural precipitation in dry years. r_t either follows the deterministic, i.i.d scenario, or the Markov chain assumption highlighted above.

Dynamic Programming Problem

Since the objective of this paper is to implement stochastic rainfall to address welfare gains from optimal management, we use a dynamic programming approach (Bellman 1957). A discussion of dynamic programming as applied to a groundwater setting can be found in Provencher & Burt (1994). The value function represents the expected present value of future benefits of the system assuming optimal management in all future periods.

$$V(x_t) = \max_{w_t} E_r \left[\sum_{t=1}^{\infty} \beta^t \pi_t(w_t, x_t | r_t) \right] \quad (10)$$

By applying the principle of dynamic programming, the first order conditions for this problem are given by the Hamilton-Jacobi-Bellman equation for this discrete time stochastic model (Brito 2007).

$$V(x_t) = \max_{w_t} \{ \pi_t(w_t, x_t | r_t) + \beta E_r [V(g(w_t, x_t | r_t))] \} \quad (11)$$

Where β is the discount factor equal to $\frac{1}{1+\delta}$, and δ the discount rate. E_r is the expectation operator over the random variable r_{t+1} and $V(g(w_t, x_t))$ the value of next period's stock assuming optimal behavior in all subsequent periods. $g(w_t, x_t)$ represents the transition equation as a function of current extraction decisions and groundwater height, which in our case is the same as equation (8) above.

Intuitively, the optimal extraction of groundwater balances the marginal benefit today against the discounted marginal benefits in all subsequent time periods given the expectations of the random variable. The first order condition with respect to withdrawals for the Bellman equation above implies the following

$$\frac{\partial \pi(w_t, x_t)}{\partial w_t} + \beta E_r \left[\frac{\partial V}{\partial x_{t+1}} \frac{\partial g(w_t, x_t)}{\partial w_t} \right] = 0 \quad (12)$$

Inputting equations (6 and 8) into equation (12),

$$\gamma(x_t)A \left[\frac{\partial F_I(w_t, x_t)}{\partial w_t} \right] + \beta E_r \left[\frac{\partial V}{\partial x_{t+1}} \left(-\frac{(1-\alpha)}{S} \right) \right] = 0 \quad (13)$$

Although we do not know V explicitly, along the optimal path, the marginal value of extraction today (the first term) should equal the discounted marginal cost (the second

term) imposed by period t 's extraction. The envelope theorem implies that the derivative of the value function with respect to the groundwater stock level generally is:

$$\frac{\partial V}{\partial x_t} = \frac{\partial \pi(w_t, x_t)}{\partial x_t} + \beta E_r \left[\frac{\partial V(g(x, w))}{\partial x_{t+1}} \frac{\partial g(w_t, x_t)}{\partial x_t} \right] \quad (14)$$

Solving for $\frac{\partial V}{\partial x_{t+1}}$ in (13) and substituting that into equation (14), moving ahead one period and rearranging terms, we arrive at the discrete time Euler equation, which does not contain the unknown value function.

$$\frac{\partial \pi}{\partial w_t} = \beta E_r \left[\frac{\partial \pi(w_{t+1}, x_{t+1})}{\partial x_{t+1}} + \frac{\frac{\partial \pi(w_{t+1}, x_{t+1})}{\partial w_{t+1}}}{\frac{\partial g(w_{t+1}, x_{t+1})}{\partial w_{t+1}}} \frac{\partial g(w_{t+1}, x_{t+1})}{\partial x_{t+1}} \right] \frac{\partial g(w_t, x_t)}{\partial w_t} \quad (15)$$

This represents the tradeoff between withdrawals today, the first term, and the opportunity cost, the part after the equal sign. With our functions,

$$\begin{aligned} \gamma(x_t) \left[\frac{\partial F_I(w_t, x_t)}{\partial w_t} \right] & \quad (16) \\ &= \beta E_r \left\{ \left[\frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}} (F_I(w_{t+1}, x_{t+1} | r_{t+1}) - F_D(r_{t+1})) \right. \right. \\ & \quad \left. \left. + \frac{\partial F_I(w_{t+1}, x_{t+1})}{\partial x_{t+1}} \gamma(x_{t+1}) \right] \right. \\ & \quad \left. + \frac{\left[\frac{\partial F_I(w_{t+1}, x_{t+1} | r_{t+1})}{\partial w_{t+1}} \right]}{\left(-\frac{(1-\alpha)}{S} \right)} \left[\gamma(x_{t+1}) + \frac{R\gamma(x_{t+1})}{AS \frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}}} \right] \right\} \left(-\frac{(1-\alpha)}{S} \right) \end{aligned}$$

The opportunity cost includes many components. The effect of a marginal reduction

of irrigated land is $\frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}} (F_I(w_t, x_t) - F_D)$, where $\frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}} > 0$. This captures the

loss at the margin of moving to dryland farming since a drop in water height decreases

irrigated land. This is multiplied by returns from irrigated agriculture as compared to dryland farming. These losses are also larger under small realizations of rainfall because the irrigated production has the ability to increase yields while dryland farming is a victim of circumstance. The value of the backstop technology plays an important role in management, since as $(F_I - F_D)$ grows, or as dryland farming returns become negligible to irrigated farming returns, the opportunity cost of extraction increases.

The next term captures the increased pumping cost on returns from irrigated agriculture $\frac{\partial F_I(w_t, x_t)}{\partial x_{t+1}} \gamma(x_t)$, scaled by how much irrigated land exists. The term, $\frac{\left[\frac{\partial F_I(w_{t+1}, x_{t+1})}{\partial w_{t+1}} \right]}{\left(-\frac{(1-\alpha)}{S} \right)}$, captures the future marginal benefit of extraction divided by the marginal change to the height of water. This is the pumping height externality when not accounted for by extractors. This is multiplied by, $\gamma(x_{t+1}) + \frac{R\gamma(x_{t+1})}{AS \frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}}}$, which is the additional marginal effect of the change in height on future extraction benefits and captures the effect on the changing irrigated acreage through $\gamma(x_{t+1})$, as well as the effects on height through recharge.

It may be helpful to compare this model to a special case; if we assume $\gamma(x) = 1$ and is constant, and $F_D = 0$, we have the bathtub model which is the extreme case where there is no spatial depletion and no backstop technology available. This case matches other models used in the literature (Gisser and Sanchez 1980, Feinermann and Knapp

1983). Using these restrictive assumptions gives us the equivalent to equation (16) for the bathtub model.

$$\left[\frac{\partial F_I(w_t, x_t)}{\partial w_t} \right] = \beta E_r \left\{ \left[\frac{\partial F_I(w_{t+1}, x_{t+1})}{\partial x_{t+1}} \right] + \frac{\left[\frac{\partial F_I(w_{t+1}, x_{t+1})}{\partial w_{t+1}} \right]}{\left(-\frac{(1-\alpha)}{S} \right)} \right\} \left(-\frac{(1-\alpha)}{S} \right) \quad (17)$$

We can see the term $\frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}} (F_I(w_{t+1}, x_{t+1}) - F_D)$, appears in equation (16) but not (17) and is positive when $F_I > F_D$, implying an additional opportunity cost of extraction over the bathtub model. The term in equation (16), $\gamma(x_{t+1}) + \frac{R\gamma(x_{t+1})}{AS \frac{\partial \gamma(x_{t+1})}{\partial x_{t+1}}}$, represents the additional impact on future benefits from the change in the water table in the spatial depletion model. If this is greater than unity it provides an additional cost in the spatial depletion model compared to the bathtub model. It is unclear if this term is always greater than one and provides an addition cost to extraction compared to the bathtub model. It is clear that when $\gamma(x_{t+1})$ is larger that there are greater costs than when $\gamma(x_{t+1})$ is smaller given assumption about the rate of depletion.⁹ However, a result of the cone assumption is that as the water height is lower in the aquifer, a similar recharge will result in a larger height change due to the reduced size of the volume it is filling. Therefore, the total effect of adding the spatial externality is a matter of parameterization.

⁹ The key assumption here would be that the second derivative of the depletion function is constant, and the rate of change in irrigable area were constant, then the costs of the change are higher when $\gamma(x)$ is relatively large because it only enters in the numerator of equation (16).

Solving equation (11) means finding the function $V(x)$ either explicitly or numerically. The addition of a stochastic element and a non-linear spatial depletion function makes analytic solution intractable in this case (Brito 2007). For the example below, we turn to the numerical method of value function iteration. Total welfare gains from management are calculated by differencing the present value of the system from the initial groundwater level to reaching the steady state under both optimal and myopic behavior. Myopic behavior is used to simulate perfect competition, as farmers would behave as if they do not expect to have the rights to the future benefits of water savings in the current period due to open access properties of the aquifer. Therefore, this behavior is simulated as solving one period's profit maximization problem, maximizing (6) with respect to withdrawals, with no regard to future states by equating marginal benefit to marginal cost for each period sequentially.

Empirical Model

We quantify welfare gains in this spatial depletion model under uncertainty with an application to the Northwest Kansas section of the Ogallala aquifer. The Kansas Water Authority provided the physical parameters of the aquifer (hydraulic conductivity, recharge, storativity). The pumping cost estimates from this region are taken from Hendricks & Peterson (2012) for groundwater management district 4 in Kansas. We estimate the crop yield response to applied water using Kansas State's Crop Yield Predictor tool for a variety of crops to estimate irrigated and dryland returns to crops under various rainfall scenarios.

The physical parameters of the model are presented in Table 1. The spatial externality is modeled using the function in equation (1), which captures the loss of irrigated acreage as groundwater levels fall. The initial aquifer area is set at 2.19 Million acres, of which 373,200 acres are irrigated or 17% of the total land. We assume that this total acreage of farmland (irrigated plus dryland) remains constant as the aquifer area is depleted. A discussion of the implications of allowing the amount of farmland to vary is found in the discussion.

The bottom of the aquifer is set at 2,892 feet above sea level based on the minimum water table found over management district 4. The initial water level was set at 3,069 feet based on the irrigated acreage and initial pumping height to the surface of 26 feet, which is the average across management district 4. The depth of the aquifer and initial total acreage surface area make up the two physical parameters needed to define the cone shaped function used to capture the spatial depletion, Figure 1. As discussed in equation (16) above, the gradient of this function describing depletion has implications to the optimal extraction time path, as it defines the magnitude of the marginal cost of lost irrigated acreage due to a change in groundwater height, or our spatial externality.

To estimate crop yields as a function of rainfall and applied irrigation water, we used Kansas State's Crop Yield Predictor tool, which is parameterized for the Colby, KS area (Klocke et al. 2010). We fit functions for corn, sorghum and winter wheat. By running the tool for the full range of water applied to crops and rainfall, we estimated

per-acre yield as a function of rainfall, or rainfall plus irrigation water applied. The assumption used here is that an inch of rainfall is equivalent to an inch of irrigation water applied. We fit cubic functions to the yield data. Visually, the yield functions are found in Figure 3. Corn is assumed grown on irrigated acreage and a rotation of wheat, fallow, sorghum, fallow on dryland acreage (Hansen et al. 2012). Thus, the crop yields from dryland are assumed to be 1/3 of the per-acre sorghum yield at a particular rainfall plus 1/3 wheat. This assumes that in any year, various dryland farms, which are represented as homogenous, are at various stages of the rotation. Prices were obtained from USDA ERS Yearbook Tables and average US prices in 2013/2014 were used in the parameterization and simulation (USDA ERS 2014).

Expectations of rainfall play an important role in defining optimal management. The model presented is flexible to various definitions of rainfall expectations and stochastic processes. For clarity of interpretation and limiting the computational burden, we chose to use three, roughly equally likely, levels of rainfall representing low, medium and high amounts of yearly rainfall¹⁰. We fit an empirical time-homogenous¹¹ Markov chain process to binned rainfall amounts to match the low, medium, and high levels to observed rainfall at the Colby, KS gauge. The transition probability matrix and details are found in Table 2.

¹⁰ Yearly rainfall is the input to the Crop Yield Predictor. From there, a weather generator is used to create growing season weather to estimate crop yield.

¹¹ This implies a stationary distribution and constant transition matrices. By simulating this process for 250,000 years, we estimated the non-conditional probabilities of each state, which are used for our stochastic case, as well as the average of the process for the deterministic treatment to match the conditions for each scenario for comparison.

Numerical Solutions

Optimal policy functions are found numerically by means of stochastic dynamic programming using MATLAB 2013. The optimal extraction decisions as a function of water level are approximated by first estimating the value function, $V(x)$, by means of value function iteration by solving equation (11) for discrete levels of x , replacing the optimized values for $V(x)$ and repeating until the functions converge within a specified tolerance (Putterman 1994). Creating the transition probability matrix for each realization of rainfall incorporates the Markov chain yearly rainfall process. Once the optimal value functions are found, the corresponding policy functions can be recovered for any realization of rainfall and groundwater height. The value function and policy function are displayed in Figure 5 and 6.

The derived optimal decision rules are iterated through time starting at the initial groundwater levels. Realizations of rainfall in each year are generated from the i.i.d or the Markov process fitted above to yearly rainfall to match expectations used for transition probabilities. To evaluate gains from optimal management, a myopic decision is made in each year by using the policy of maximizing equation (6) for each year with respect to groundwater withdrawals. These paths through time for the i.i.d case can be found in Figure 6.

The welfare implications of making optimal extraction decisions are estimated by discounting and summing each year's profit as defined by equation (6) from the initial

year to the end of the time horizon, in our case 500 years, which is well after reaching a steady state and where future benefits are essentially discounted to zero. The differences between the two welfare values calculated are the estimated gains from optimal management of the aquifer. The results are presented in Table 3 for different specifications of the problem.

Results

We present results from our model applied to the characteristics of groundwater management district 4. The welfare results can be found in Table 3 under the various rainfall assumptions. Gains exist from management in the range of 2.88-3.01%. The stochastic scenarios generate slightly larger relative gains (.09-.13% larger), reflecting the magnitude of including stochasticity in assessing welfare impacts to management. A point found in the literature, reflecting the additional value of groundwater found by previous studies under uncertainty (Knapp & Olson 1995; Tsur & Graham-Tomasi 1991; Provencher & Burt 1993).

The results show that including the loss of the spatial extent of the aquifer leads to a larger estimated welfare losses when compared to other modeling assumptions for welfare impact studies over parts of the Ogallala. To put our work into perspective, our estimates are larger than Lee et al (1981) for the Ogallala in Texas at .3%, Nieswiadomy (1985) for the High plains aquifer in Texas at .28%, and similar to Kim et al. (1989) also for the High plains in Texas (1-3.7%). Kim's paper includes

endogenous technological change as a function of groundwater heights and therefore similarly provides backstop technologies in a sense, albeit under certainty.

The welfare gains arise from the difference between the optimal decision rules, and the myopic pumpers' choices, Figure 5. The optimal planner internalizes the inter-temporal stock externality, while the myopic extractor does not. For the stochastic rainfall scenarios, the realized rainfall determines the corresponding optimal policy function, one for each state of rainfall. As compared to the myopic decision maker, withdrawals following optimal management are about 16% less in low rainfall, 19% less in medium rainfall, and 25% less in high rainfall, varying through groundwater heights (Figure 5) showing relatively larger savings in better years. When these optimal policy functions are iterated through time (Figure 6) meaning the decisions are made optimally based on the current rainfall and groundwater state, the groundwater stock is depleted more slowly under optimal decisions, as expected.

The lower optimal extractions, in equivalent groundwater heights and rainfall amounts compared to the myopic pumper, leads the groundwater levels to settle at a steady state height of 2,991 feet on average as opposed to 2,980 feet under myopic extractions for a difference of 11 feet in height at the steady state and between 0 and 17.5 feet through the time-path. With our spatial cone model, where the irrigated acreage is a function of the groundwater height, these steady state groundwater levels translate to an optimal steady state with 118,360 irrigated acres or 31% of the initial irrigated acreage, as opposed to 93,517 irrigated acres under myopic extraction or 25% of the initial irrigated acreage. The sources of welfare losses are the higher pumping

costs with deeper water tables and the loss of irrigated acres. The magnitude of these losses is less obvious as they depend on the rate of depletion, the discount rate, and the relative benefits of irrigated farmland as compared to dryland farming.

When comparing welfare across rainfall scenarios, modeling uncertainty increases estimated gains from management, resulting in an increase in welfare gains of .09-.13% over the deterministic welfare gains. The concavity of the one year benefit function implies risk aversion in the sense that irrigation water limits the range of available water in future years, which is preferred over a wider range of possible states as described in Provencher & Burt (1993) and under risk aversion in Knapp & Olson (1996). However, the yield of the backstop technology of dryland farming is less variable (flatter in Figure 3) than irrigated crop yields. In a sense, the transition to dryland farming attenuates the loss of the buffer value of the groundwater, as the crops are more resilient to a range of rainfall than non-irrigated corn would be. Overall, the increase in welfare gains when including uncertainty reflects the additional benefit gained when taking into account groundwater as a steady resource to buffer variation in weather in optimal management.

The groundwater height remains higher through time when taking into account uncertainty in precipitation as the optimal policy consists of slightly smaller extractions on average. The average difference in height is small and only around 1 foot, well within the margin of error of the simulation runs. Clearly, the year-to-year extractions are different in the stochastic cases from the deterministic case, where

optimal policy functions imply changing policy rules based on yearly rainfall, Figure 5. So, although the extraction paths are similar in average groundwater height, the stochastic case leads to variable extractions and a range of groundwater levels depending on the realization of rainfall (Figure 6). The addition of a Markov chain process added little additional welfare gains (.04%) when optimal policy was matched to a stochastic process as described in Table 2.

Discussion

Unlike other approaches, our spatial depletion model captures the gradual loss of access to a stabilizing resource as a result of the overall extraction of the aquifer. The imbedded assumption of infinite conductivity results in uniform pumping heights across the irrigated acreage above the remaining aquifer, representing the average height faced by the remaining pumpers. The intertemporal pumping height externality, assuming infinite conductivity, is relatively small when the marginal pumping cost as a function of height is low compared to the marginal benefit of groundwater extraction (Koundouri 2004). Thus, the important externality to welfare is the loss of access over time as certain areas have thinner saturated thickness than others and, even with a relatively short pumping height, can lose access as a result of the aquifer wide extractions, Figure 7. When simulating the aquifer under the assumption of the standard bathtub model, the estimated welfare gains are small, .06% (Appendix Table 1).

That modeling the loss of irrigated acreage does not result in significantly larger welfare losses than what we found was unexpected, although our welfare estimates are larger than those found over similar regions in previous literature. We include a backstop technology, dryland farming, which reduces the gains from management, since in its absence, we would have assumed infinitely expensive alternatives (Koundouri & Christou 2006). Without a backstop, the welfare gains from management are larger at 5.2% in the stochastic rainfall scenario (Appendix Table 1). We chose to model the dryland crop as a rotation of sorghum, wheat, and fallow in equal proportions. Dryland practices require fallowing some areas and rotations of other crops to maintain soil conditions. We assumed no loss of total farmland, when some land may have to be fallowed permanently after losing access due to its soil characteristics, which may cause us to overstate the value of the backstop technology. Advances in dryland techniques and drought resistant crops could mitigate some of the negative consequences to losing irrigation access, but even as dryland techniques gain efficiency, so too will the returns to irrigated agriculture and it is unclear if the relative value difference will change much (Steward et al. 2013).

Second, our model assumes that 17% of the initial aquifer surface area is irrigated farmland (Table 1). This is based on the observed number of acres farmed over this district. The total surface area of the usable aquifer, or initial irrigation intensity, may vary widely across aquifers or regions within an aquifer. The results are particularly sensitive to the intensity of farming, or the amount of farming relative to the size of the resource, with a higher percent of acres irrigated leading to larger welfare gains

from management (Appendix Figure 1). This suggests that intensely farmed areas or aquifers may have a magnitude of difference in gains from optimal management.

In exploring the implication of the above assumptions, we can say something about when management is likely to have a greater welfare gain. When the value of the backstop technology is small compared to the value of the use with access to the water, the welfare gains are larger. Koundouri and Christou (2006) made this point with a backstop of desalinization under certainty and we can see this analytically in our model and discussion of equation (16). Under uncertainty, the backstop technology's benefit of reducing variability also plays a role given the concavity of the benefit function implying risk aversion. A less variable backstop may have more value than one of similar average value but more variable returns and result in smaller gains to management, although this is not proven here.

We also held prices constant through the time horizon, but adding demand growth for the irrigated crop would certainly increase the welfare gains, similar to what is shown in Brill & Bruness (1994). There may be important price dynamics in large aquifers that support large monocultures, as irrigation water dries up the prices of irrigated crops should increase and the opposite market forces would be in play for dryland crops. These forces would increase the gains from management as the marginal benefits from irrigated agriculture would increase compared to dryland agriculture.

Our motivation in including uncertainty was that groundwater has a value as a buffer to smooth out returns over time (Tsur & Graham-Tomasi 1991; Knapp & Olson 1995). With a backstop that does not enjoy the stable source of groundwater, we surmised the two aspects together could lead to additional welfare gains from moving away from myopic pumping. Despite the small additional percentage gains from management, a few points can be made from interpreting the optimal rules under uncertainty. The optimal extractions vary year to year, so any management designs to approximate the optimal extractions would need to allow for this variability as rules based on a deterministic optimal solution would limit welfare beneficial pumping in dry years and similarly be too high to induce optimal saving in better years (Figure 5).

The groundwater stock paths are similar under uncertainty and certainty on average, but vary in the stochastic cases, shown Figure 6. The gains from management could be small and similar to the deterministic case or larger than the average depending on the realization of rainfall, only one of which the resource would actually progress through in reality.

We included a Markov process to investigate if the type of stochastic rainfall process mattered to welfare gains to management, as this is not something that has been discussed to our knowledge. We find that changing the process mattered little to the estimated gains from management, although it changed the optimal policy rules. In our case it slightly increased extractions in each rainfall state, in each groundwater level although the differences are well within the modeling uncertainty. Given that adding

uncertainty in general added less than a quarter percent to welfare gains, finer distinctions, such as defining a different process, did not materially matter.

In our model, we simplified droughts with a Markov chain rainfall process where the lengths of the droughts were defined by the transition probabilities. Drought refers to extended periods of lower than normal soil moisture as a result of climatic variables including rainfall (used in this paper), temperature, and even wind speed. Drought measures such as the Palmer Drought Severity Index include a duration aspect, where the progressions of the climatic variables matter and not just the current conditions. We do not carry over soil conditions or other stock variables, except the groundwater height, that could be affected by the series of yearly rainfall events. This could be done by modeling additional processes in the benefit function and is left for future work. The addition of the lagged process mattered little to the welfare gains from management and only slightly to the optimal policy functions in our application.

The discussion of rainfall processes leads to an important point about the assumptions leading to the optimal paths. We matched the rainfall processes to the optimal policy rules in each case assuming perfect information about the process that generates the annual rainfall. Any policy rules would be such that the process is not known with certainty. The method we chose to use to estimate optimal management assumes that the decision makers not only act optimally in each period given the realization of rainfall, but also have accurate expectations of the process that generates rainfall. In reality, the processes that generate weather and longer-term climate trends are

complex and the distribution of rainfall is not stationary. As such, the extent to which the policy maker's expectation of rainfall coincides with the actual process may have implications to welfare gains, with mistakes or mis-specifications causing possibly large welfare losses.

Conclusion

We introduce a dynamic spatial depletion model of groundwater extraction that incorporates stochastic rainfall and a gradual spatial stock externality, leaving more farmland at the mercy of variable rainfall as groundwater levels fall and less available for irrigation across a smaller area. By building a novel model flexible to rainfall expectations and various stochastic processes, we showed the extent of the importance to welfare and optimal management when including stochastic elements. We also explore the implication of droughts and their persistence on optimal management as it compares to myopic extraction behavior. We find that the addition of randomness and persistence of rainfall does not materially affect welfare gains, largely due to the relatively good yields from dryland farming. Incorporating the spatial depletion of the aquifer added to welfare estimates of moving from open access to optimal management of the resource.

REFERENCES

- Bellman, R., 1957. *A Markovian decision process*, DTIC Document. Available at: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=A0606367> [Accessed January 24, 2014].
- Bill Provencher & Oscar Burt, 1994. Approximating the Optimal Groundwater Pumping Policy in a Muliaquifer Stochastic Conjunctive Use Setting. *Water Resources Research*, 30(3), pp.833–843.
- Brill, T. & Bruness, S., 1994. Planning Versus Competitive Rates of Groundwater Pumping.
- Brito, P., 2007a. Introduction to Dynamic Programming Applied to Economics. Available at: http://www.fep.up.pt/docentes/joao/material/aea/notas_pbrito_2008.pdf [Accessed January 24, 2014].
- Brito, P., 2007b. Introduction to Dynamic Programming Applied to Economics. Available at: http://www.fep.up.pt/docentes/joao/material/aea/notas_pbrito_2008.pdf [Accessed May 13, 2014].
- Brozović, N., Sunding, D.L. & Zilberman, D., 2010. On the spatial nature of the groundwater pumping externality. *Resource and Energy Economics*, 32(2), pp.154–164.
- Buchanan, R.C., Buddemeier, R.R. & Wilson, B.B., 2009. Kansas Geological Survey. Available at: <http://www.kansasforchange.org/Docs/The-High-Plains-Aquifer.pdf> [Accessed August 26, 2014].
- Feinerman, E. & Knapp, K.C., 1983. Benefits from Groundwater Management: Magnitude, Sensitivity, and Distribution. *American Journal of Agricultural Economics*, 65(4), p.703.
- Gisser, M. & Sanchez, D.A., 1980. Competition versus optimal control in groundwater pumping. *Water Resources Research*, 16(4), pp.638–642.
- Guilfoos, T. et al., 2013. Groundwater management: The effect of water flows on welfare gains. *Ecological Economics*, 95, pp.31–40.
- Hansen, N.C. et al., 2012. Research achievements and adoption of no-till, dryland cropping in the semi-arid U.S. Great Plains. *Field Crops Research*, 132, pp.196–203.
- Harou, J.J. et al., 2010. Economic consequences of optimized water management for a prolonged, severe drought in California: Economic Consequences of Prolonged Severe Drought. *Water Resources Research*, 46(5).

- Hendricks, N.P. & Peterson, J.M., 2012. Fixed effects estimation of the intensive and extensive margins of irrigation water demand. Available at: <http://krex.k-state.edu/dspace/handle/2097/13952> [Accessed January 29, 2014].
- Hornbeck, R. & Keskin, P., 2012. *Does Agriculture Generate Local Economic Spillovers? Short-run and Long-run Evidence from the Ogallala Aquifer*, National Bureau of Economic Research. Available at: <http://www.nber.org/papers/w18416> [Accessed January 24, 2014].
- Katic, P.G., 2010. *Spatial dynamics and optimal resource extraction*, Centre for Water Economics, Environment and Policy, Crawford School of Public Policy, The Australian National University. Available at: http://www.cer.ethz.ch/sured_2010/programme/SURED-10_051_Katic.pdf [Accessed May 29, 2014].
- Kim, C.S., Moore, R. & Hanchar, J.J., 1989. A dynamic model of adaptation to resource depletion: Theory and an application to groundwater mining.
- Klocke, N.L. et al., 2010. Scheduling for Deficit Irrigation-Crop Predictor. *Applied Engineering in Agriculture*, 26(3), pp.413–418.
- Knapp, K.C. & Olson, L.J., 1996. Dynamic resource management: Intertemporal substitution and risk aversion. *American journal of agricultural economics*, pp.1004–1014.
- Knapp, K.C. & Olson, L.J., 1995. The economics of conjunctive groundwater management with stochastic surface supplies. *Journal of Environmental Economics and Management*, 28(3), pp.340–356.
- Koundouri, P., 2004. Current issues in the economics of groundwater resource management. *Journal of Economic Surveys*, 18(5), pp.703–740.
- Koundouri, P. & Christou, C., 2006. Dynamic adaptation to resource scarcity and backstop availability: theory and application to groundwater*. *The Australian Journal of Agricultural and Resource Economics*, 50(2), pp.227–245.
- Lee, K., Short, C. & Heady, E., 1981. Optimal Groundwater Mining in the Ogallala Aquifer: Estimation of Economic Losses and Excessive Depletion Due to Commonality. Available at: <https://ideas.repec.org/p/ias/cpaper/81-wp1.html#biblio>.
- Leizarowitz, A. & Tsur, Y., 2012. Renewable resource management with stochastic recharge and environmental threats. *Journal of Economic Dynamics and Control*, 36(5), pp.736–753.
- McGuire, V.L., 2012. Water-level and storage changes in the High Plains aquifer, predevelopment to 2011 and 2009–11. Available at: <http://digitalcommons.unl.edu/usgspubs/123/> [Accessed May 22, 2014].

- Nieswiadomy, M., 1985. Demand for irrigation water in the High Plains of Texas, 1957-80.
- Palazzo, A. & Brozović, N., 2014. The role of groundwater trading in spatial water management. *Agricultural Water Management*, 145, pp.50–60.
- Provencher, B. & Burt, O., 1993. The Externalities Associated with the Common Property Exploitation of Groundwater. *Journal of Environmental Economics and Management*, 24(2), pp.139–158.
- Putterman, M.L., 1994. Markov decision processes. *John Wiley and Sons, New York*.
- Roseta-Palma, C. & Xepapadeas, A., 2004. Robust control in water management. *Journal of Risk and Uncertainty*, 29(1), pp.21–34.
- Saak, A.E. & Peterson, J.M., 2012. Groundwater pumping by heterogeneous users. *Hydrogeology Journal*, 20(5), pp.835–849.
- Saak, A.E. & Peterson, J.M., 2007. Groundwater use under incomplete information. *Journal of Environmental Economics and Management*, 54(2), pp.214–228.
- Savage, J. & Brozovic, N., 2011. Spatial externalities and strategic behavior in groundwater pumping. In *AERE Summer Conference*. Available at: http://webmeets.com/files/papers/AERE/2011/188/savage_gw_5_21_11.pdf [Accessed May 22, 2014].
- Srikanthan, R. & McMahon, T.A., 1999. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, 5(4), pp.653–670.
- Steward, D.R. et al., 2013. Tapping unsustainable groundwater stores for agricultural production in the High Plains Aquifer of Kansas, projections to 2110. *Proceedings of the National Academy of Sciences*, 110(37), pp.E3477–E3486.
- The Economist, 2013. Water and Agriculture in Kansas: Sip it Slowly. *The Economist*.
- Tsur, Y. & Graham-Tomasi, T., 1991. The buffer value of groundwater with stochastic surface water supplies. *Journal of Environmental Economics and Management*, 21(3), pp.201–224.
- Tsur, Y. & Zemel, A., 2004. Endangered aquifers: Groundwater management under threats of catastrophic events: Endangered Aquifers. *Water Resources Research*, 40(6).
- United Nations Environment Programme, 2012. A Glass Half Empty: Regions at Risk Due to Groundwater Depletion. Available at: http://na.unep.net/geas/getuneppagewitharticleidsript.php?article_id=76 [Accessed January 14, 2015].

- United Nations Environment Programme & Division of Early Warning and Assessment, 2003. *Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management*, Nairobi, Kenya: UNEP.
- USDA ERS, 2014. Feed Grain: Yearbook Tables. Available at: <http://www.ers.usda.gov/data-products/feed-grains-database/feed-grains-yearbook-tables.aspx#.U-ys4lZnv6E> [Accessed August 14, 2014].
- Wijnen, M. et al., 2012. *Managing the Invisible. Understanding and Improving Groundwater Governance*, World Bank Publications.
- Zeitouni, N., 2004. Optimal extraction from a renewable groundwater aquifer with stochastic recharge. *Water Resources Research*, 40(6).

Tables and Figures

TABLE 1
Parameter values for Northwest Kansas Groundwater Management District 4

Parameter	Description	Value
C_1	Cost of pumping	\$.11 /a-ft/ft
R	Natural recharge	199,040 a-ft
A	Aquifer area	2.19 million acres
	Initial Irrigated acres	373,200 acres
\bar{X}	Land surface	3094 ft above sea-level
\underline{X}	Lower aquifer bound	2892 ft above sea-level
S	Storativity	.17
α	Irrigation water return	20%
x_0	Initial water level	3068 ft above sea-level
β	Discount factor	.96%
r	Rainfall states	
	High	2 ft
	Medium	1.58 ft
	Low	1.25 ft
p	Crop Prices	
	Corn	\$4.45 /bushel
	Sorghum	\$4.25 /bushel
	Winter Wheat	\$6.53 /bushel

Note- The physical parameters of the aquifer are provided by the Kansas Water Authority. The pumping cost estimates from this region are taken from Hendricks & Peterson (2012) for groundwater management district 4 in Kansas. Prices were obtained from USDA ERS Yearbook Tables and average US prices in 2013/2014 were used in the parameterization and simulation (USDA ERS 2014).

TABLE 2
Markov Chain Transition Probability Matrix

Future State	Current State			Non-conditional
	Low	Med	High	
Low – 1.25 ft	40.48%	21.62%	42.50%	35.29 %
Med – 1.58 ft	30.95%	37.84%	25.00%	31.10 %
High - 2 ft	28.57%	40.54%	32.50%	33.61 %

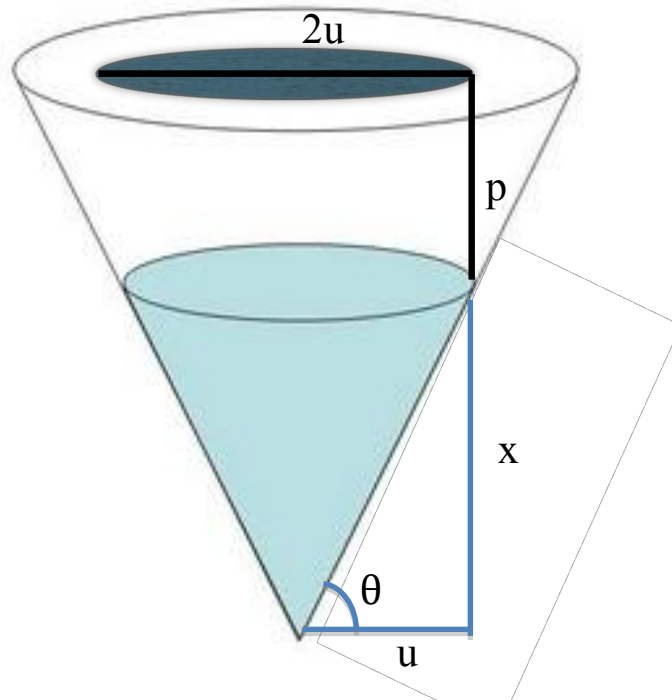
NOTE- Low Med and High refer to the annual precipitation events. Probabilities are empirically found from a precipitation time series from Colby, KS. The last column represents the non-conditional (on current state) probabilities of future rainfall used in the i.i.d. stochastic scenario.

TABLE 3
Welfare Gains From Groundwater Management
Total Discounted Profit (Billion \$)

	Perfect Competition	Optimal Policy	Difference	% Gain
Deterministic	\$ 6.683	\$ 6.876	\$.193	2.88
Stochastic	\$ 6.589 (.0937)	\$ 6.785 (.0938)	\$.196 (.006)	2.97 (.04)
Stochastic -MC	\$ 6.515 (.0408)	\$6.711 (.0350)	\$.196 (.006)	3.01 (.03)

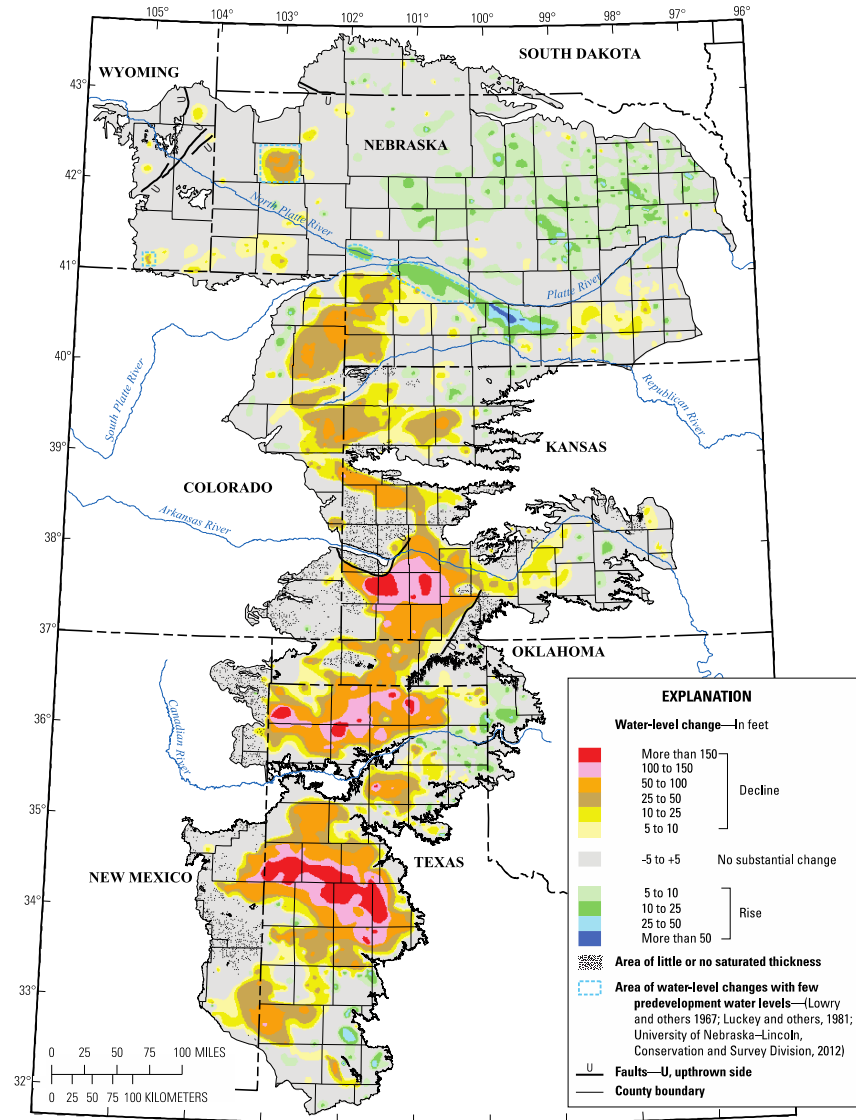
NOTE- Standard errors of the stochastic figures from 100 iterations through rainfall realizations are in parentheses. The deterministic scenario assumes average annual rainfall each year. Stochastic assumes i.i.d. random draws from high, average, low rainfall state based on empirical probabilities. Stochastic- MC assumes draws from a Markov chain process where the transition probabilities are found in TABLE 2.

FIGURE 1
Spatial Depletion Model



NOTE – u is the radius of the irrigated acreage, x the groundwater height and p is the pumping height, which is the difference between the surface and the height of the groundwater. θ is fixed given the parameters x , p , and u .

FIGURE 2
Water Level and Storage Change in the High Plains Aquifer
Predevelopment (1950) to 2011

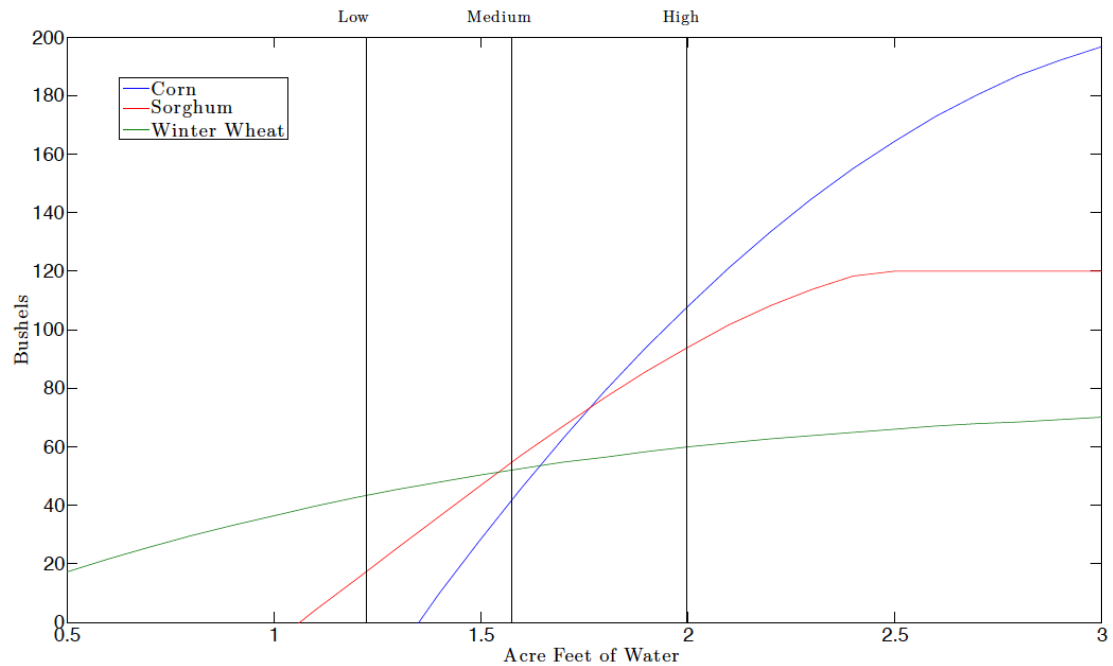


Base from U.S. Geological Survey digital data, 2001, 1:2,000,000
 Albers Equal-Area Conic projection
 Standard parallels 29° 30' and 45° 30', central meridian -101°
 North American Datum of 1983 (NAD 83)

Aquifer boundary from Qi (2010); areas of little or no saturated thickness and faults from Gutentag and others (1984) and Cederstrand and Becker (1999a, 1999b)

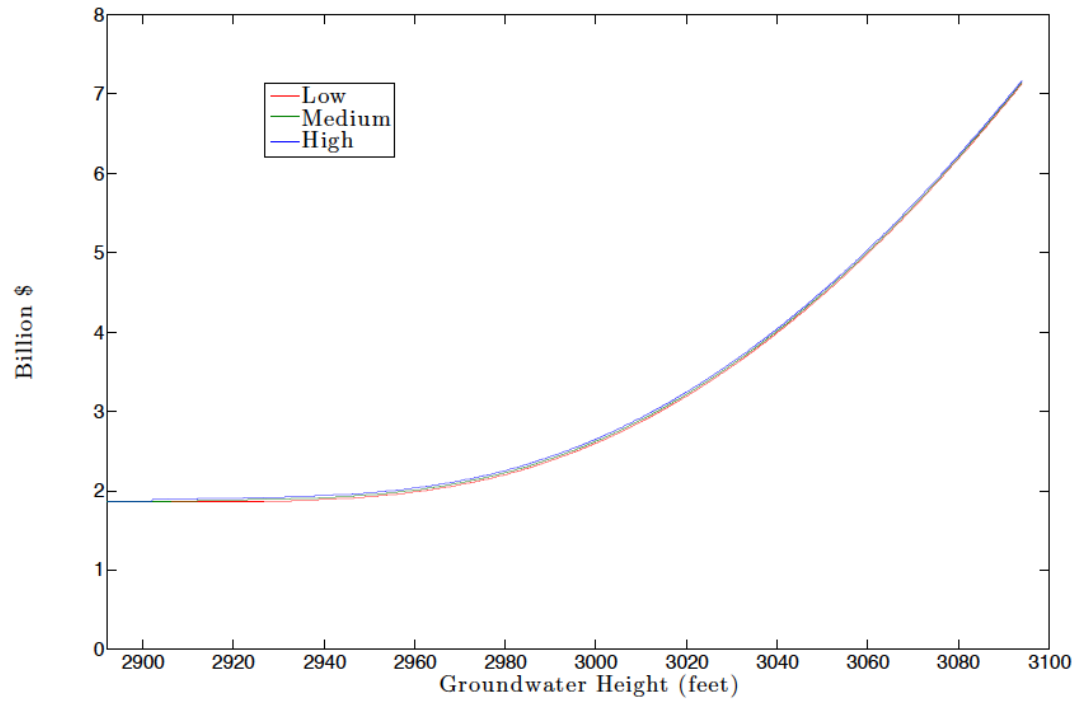
(McGuire 2012)

FIGURE 3
Crop Yield Functions
Bushels per-acre



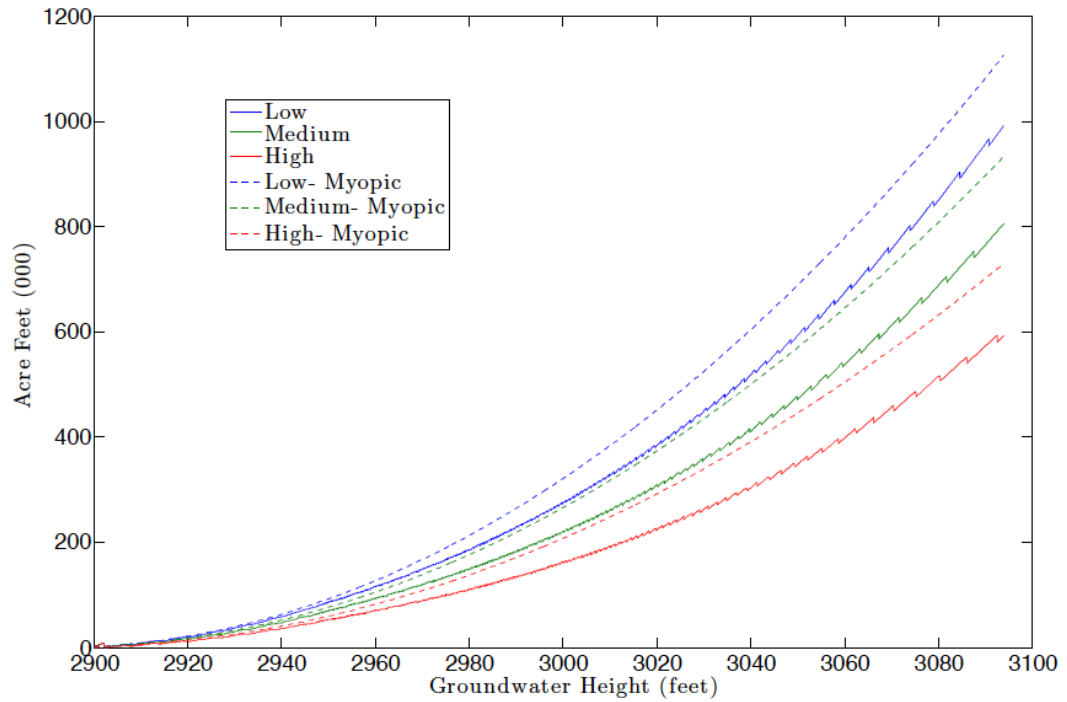
NOTE- Derived from Kansas State University's Crop Yield Predictor. Low, medium and high rainfall amounts are shown as vertical lines. The Y-axis represents yield in terms of bushels per-acre as a function of rainfall for sorghum and winter wheat, or rainfall plus irrigation water for corn.

FIGURE 4
Optimal Value Functions



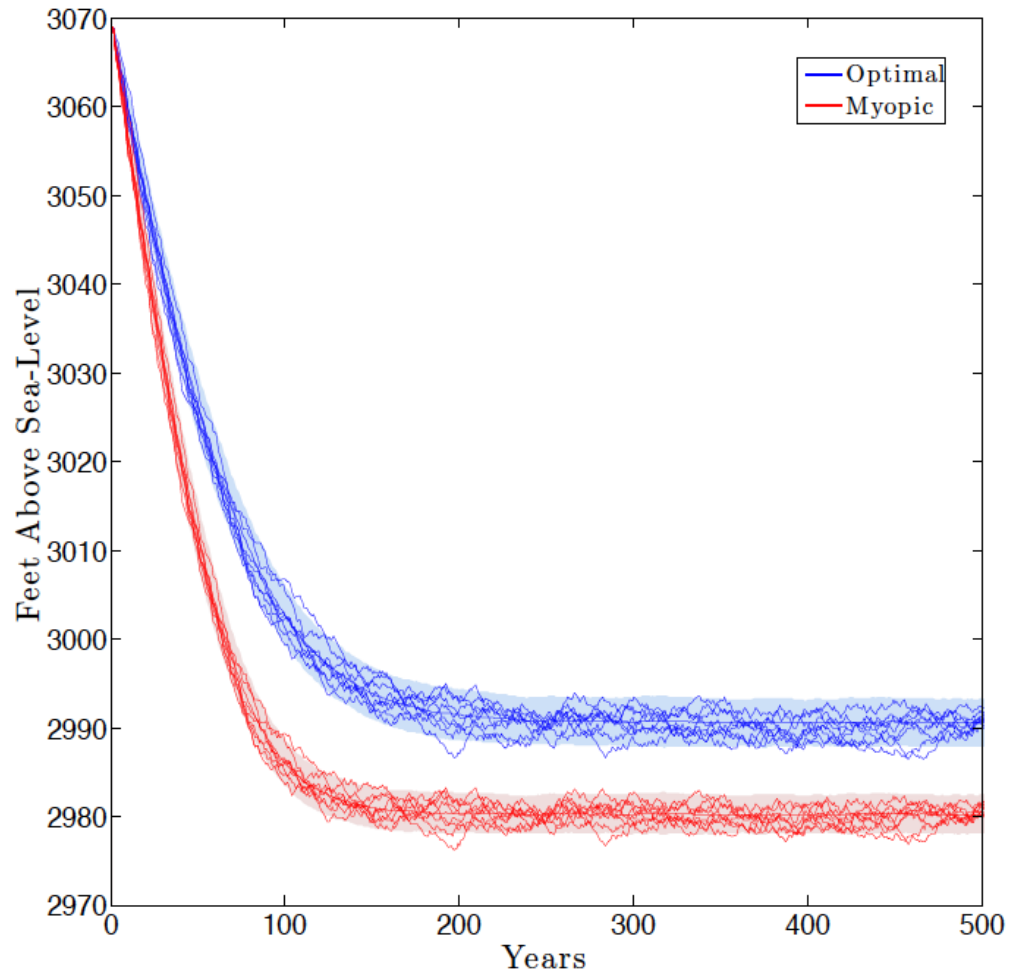
NOTE- This figure plots the optimal value functions for the stochastic (one for each realization of rainfall). The value function represents the net present value of the groundwater resource assuming optimal future management.

FIGURE 5
Optimal Policy Functions at Groundwater Heights
i.i.d. Rainfall



NOTE- The three states of rainfall (Low, Medium, High) each has a corresponding policy function represented by the three colors. At low levels of rainfall, demand for irrigation water is higher. The dotted lines represent the decisions made by a myopic extractor, while the solid lines are results of the optimal dynamic programming policy functions.

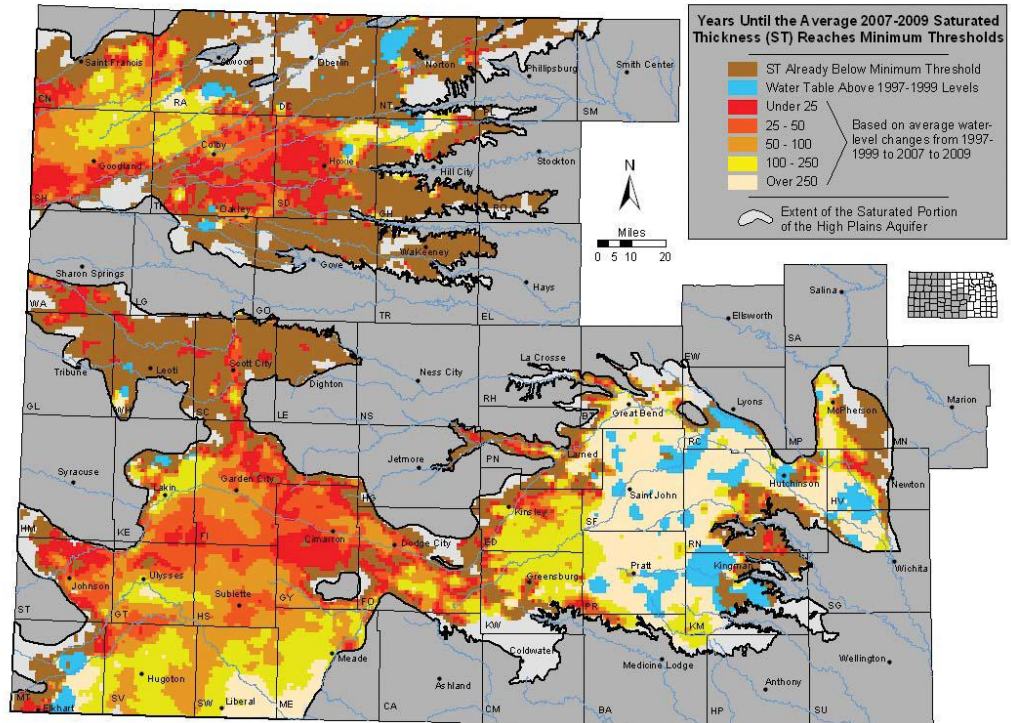
FIGURE 6
Groundwater Height Over Time
i.i.d Rainfall



NOTE- The blue lines represent groundwater levels under optimal management. The average (thicker line), standard deviation (blue shaded region) and a number of individual runs are plotted. The red is the myopic planner's groundwater heights through time.

FIGURE 7

Estimated Usable Lifetime of Ogallala in Kansas



NOTE- From Kansas Geological Survey (2009). The 1998-2008 trends of depletion were extrapolated to estimate useable lifespan.

Manuscript 3

Weather, Energy and Capacity: Dynamics of the Natural Gas Market in New England

By

Nathaniel H. Merrill

Department of Environmental and Natural Resource Economics

University of Rhode Island

Weather, Energy and Capacity: Dynamics of the Natural Gas Market in New England

Nathaniel H. Merrill

University of Rhode Island, Email: nateme16@my.uri.edu
Department of Environmental and Natural Resource Economics
University of Rhode Island
Coastal Institute
1 Greenhouse Rd, Kingston RI 02881

Abstract

This paper empirically estimates the effect of an increase in natural gas pipeline capacity in New England on monthly equilibrium natural gas prices and quantities for the electric energy sector. The cost of natural gas has important consequences to the wellbeing and cost of living for millions of customers either relying directly on natural gas for heating, or electric energy consumers indirectly. I present results of reduced form price and quantity regressions using Generalized Least Squares (GLS) followed by results of a dynamic simultaneous equation model (SEM) of the market system. Using derived empirical relationships, prices and quantities for natural gas are estimated under various weather scenarios as well as under current and expanded capacity to highlight the role capacity has in effecting the variability of the price of energy to the region.

Introduction

Natural gas is increasingly being used for electric energy power generation across the United States. This trend is driven by increased domestic production from shale gas as well as regional environmental initiatives. For the Northeastern United States, it has meant a shift from oil and coal-fired generation towards natural gas. Since 2001, annual power generation from natural gas has increased from 17% of total MWh produced to as much as 36% in 2015 (EIA 2013; EIA 2014). Depending on the season, this increase in demand from electric energy generators is competing with other uses for natural gas in the Northeast, such as residential heating and industrial consumption. While demand has increased across all sectors, delivery capacity for natural gas is at times constrained, especially in New England (CT, ME, MA, NH, RI, VT) with nearly all the physical capacity under contract in certain winter months and available supplies for variable or unexpected demand, such as in especially cold winter months, only available at high prices (DOE 2013). Because they are particularly reliant on the share of capacity that is not under firm contract, natural gas generators are exposed to volatile fuel costs at the same times when electric power demands are high. In turn, the electricity market passes these fuel costs on to electric utilities and from there to electric power consumers.

The seasonal scarcity issue is being addressed with new capacity in the form of new pipelines and pipeline upgrades, some of which are already in use and a number of which are planned to come online over the coming years (EIA 2013). However, the pipelines are controversial both for their impact on the planned routes and the implicit

support of the controversial production process of hydraulic fracturing, as is evident from recent news reports (Marc Levy 2014; Jay Fitzgerald 2014; Ailworth 2013).

This paper will empirically estimate the effect of an increase in natural gas pipeline capacity in New England on monthly equilibrium natural gas prices and quantities for the electric sector. Weather plays an important role in defining the demand for natural gas, due to natural gas's use for heating and electricity generation in the winter and through electricity demand for cooling in the summer. The cost of natural gas has important consequences to the wellbeing and cost of living for millions of customers either relying directly on natural gas for heating, or indirectly through electric energy consumption. This paper presents results of reduced form price and quantity regressions using Generalized Least Squares (GLS), followed by results of a dynamic simultaneous equation model (SEM) of the market system. Using derived empirical relationships, prices and quantities for natural gas will be estimated under various weather scenarios as well as under current and expanded capacity to highlight the role capacity has in affecting the variability of the price of energy to the region. This work adds to the literature by providing empirical evidence and the quantification of the effect of constrained pipeline supply in an important energy market, where weather conditions, multiple demand sectors and alternative fuels determine the cost of energy.

Background

Natural gas in New England is used by the residential, commercial, industrial and electric sectors. Total yearly consumption has increased from 710,000 to 868,000

MMcf from 2001 to 2014, with the largest percent and total volume gains from electric energy generation (Figure 1). This increased consumption came partly as a result of decreasing prices from a plentiful supply of domestic natural gas due to the technological advances in extraction using hydraulic fracturing and horizontal well drilling (DOE 2013; DOE 2015). For power generation, this has meant the addition of natural gas fired turbines to replace retiring coal and nuclear facilities. Figure 2 shows electric generation by fuel type for New England, showing natural gas replacing coal over the time frame. The total energy generation has remained mostly constant.

The demand for natural gas is highly seasonal and closely correlated with heating and cooling needs. The overall quantity consumed across all sectors peaks in the winter months. However, the electric sector follows a different seasonal pattern. The residential, commercial and industrial sectors use natural gas in higher volumes in the winter for heating, while the electric sector uses more in the summer for meeting cooling demand and less in the winter due to higher spot prices and fuel oil as an available substitute for many generators (Figure 1).

Pipeline and liquefied natural gas (LNG) infrastructure must be sized to meet the needs of the peak demand in the winter, and therefore runs below capacity for the other parts of the year. New England receives natural gas through pipelines that carry it from producing regions in Canada and states to the west such as Tennessee and Pennsylvania. The major pipelines are the Algonquin Gas Transmission, which connects to supplies coming from the South, the Tennessee Gas Pipeline, which runs

from the Gulf of Mexico through the producing regions of Pennsylvania and Ohio, and the Maritimes and Northeast pipeline from Canada. In addition, imports of liquefied natural gas (LNG) from overseas come through facilities located around Boston. As there are no major storage sites in New England, the area relies on storage and outflows from the states to the west and LNG imports to meet unexpected demand.

Pipeline customers can be divided into two types: those with firm delivery contracts for a specified capacity, and those that have interruptible contracts for pipeline capacity and the associated gas volumes. Residential and many commercial customers are serviced by local distribution companies (LDCs), utilities who have long-term contracts with the pipeline for capacity to guarantee supply to their customers (DOE 2014). To a lesser extent, industrial users also enter into longer-term agreements for supply, resulting in higher average but less volatile prices. Electric energy generators, on the other hand, generally do not enter into long-term contracts. Their supply is interruptible for a number of regulatory as well as economic reasons explained below. Therefore, the price the generators pay for natural gas varies greatly and a stable supply is at risk in times of high constraint in the system. The prices across sectors can be seen in Figure 3.

Pipeline owners operate in a highly regulated market, where the price charged for capacity is capped by the Federal Energy Regulatory Committee (FERC) in order to counteract the monopoly power and protect customers (EIA 2015a). However, firm capacity contract holders can sell their owned capacity and volume of gas in an

unregulated secondary market. These transactions are made between buyer and seller and are not cleared through a market entity. As laid out in Oliver and Mason, et al. (2014) and Oliver and Finnoff, et al. (2014), capacity constraints lead to higher spot prices in constrained demand regions and higher price spreads between pipeline nodes. Therefore, additional rents are made available for firm capacity contract holders with the existence of the secondary market.

In New England, these firm contract holders are the utilities and large industrial users. Why electric generators have not sought more firm supplies is unclear. Ideas put forward include the availability of alternative fuels for generation (oil), the inability to store gas or electric energy, the ability to pass the higher spot costs along in liberalized electric energy market, the lower average price received from more interruptible contracts in the summer, and issues of counterparty risk due to the heterogeneity in sizes of gas producers and electric energy consumers (Morris 2013; ISO New England 2015). Generators may enter more long-term agreements over time, but it is currently clear that the electric energy sector pays significantly higher prices for natural gas in times of constraint due to the market structure described above.

Constraints are prominent in the winter months when heating needs and natural gas for generating electricity compete for relatively scarce capacity. The severity of the winter in terms of low temperatures determines to what extent the pipeline constraint is binding. Particularly cold winter seasons, when combined with constrained supply, have led to high prices in the recent past, particularly the winters of 2012/2013 and

2013/2014. Monthly average prices for natural gas in the electric sector were as high as 18.52 \$/Mcf in the winter for 2012/13 and 21.95 \$/Mcf in the winter of 2013/14, when the long-term average is 6.65 \$/Mcf overall and between 8 and 9 \$/Mcf in December, January and February. Historically it is December, January and February when the constraints are the highest. The exposure of the region's energy costs to risks from colder than average winters influences both the cost of living and the cost of doing business. Using empirical relationships fit to the years 2002-2014 for natural gas prices paid by the electric sector, a simulation is created to estimate price and consumption as a function of planned pipeline capacity improvements for New England under a range of possible winter temperatures.

There is a vast literature of modeling approaches to estimate natural gas price relationships across a multitude of markets. Early approaches focused on end users' costs and substitutability between energy sources using panel and system of equation approaches to identify the elasticities of interest. A review of this early work is presented in Al-Sahlawi (1989). In that time, demand was inelastic over the short term, yet elastic over the longer run, and closely related to the price of substitutes, namely oil. Given the simultaneity and endogeneity issues in estimating price elasticities in a supply and demand system, simultaneous equation models (SEM) and variations of instrumental variable approaches were used for identification of price elasticity and fuel substitution effects (Zellner & Palm 1974).

While these methods lead to a deeper understanding of the long-term drivers of natural gas prices, fundamental factors of supply and demand, such as weather and storage, which can be important at a national and certainly a regional level were largely attributed to unexplained noise. Mu (2007), used a reduced form approach to model equilibrium prices for natural gas contracts based on the Henry Hub spot price, to show that both weather and storage deviations from normal levels are important determinants in both the average price as well as in explaining changes in daily price volatility. Both changes in weather and surprise weekly storage announcements were important in explaining the price dynamics of this particular natural gas price. Since many financial derivative contracts are based on the Henry Hub price, the work is relevant for that financial market, but modeling these financial derivative prices in Chicago tells us little about the prices for the final consumers of natural gas and little about the prices for the New England region.

Endogeneity between prices, quantities and other fundamental factors, such as gas storage, makes estimating the structural parameters of demand and supply tricky. Time series methods have been employed, including variations of the ARIMA models or VAR (vector autoregressive) models for the case of multiple time series when the equilibrium price and quantity are the variables of economic interest and not necessarily estimates of structural parameters. A number of papers (Nick & Thoenes 2013; Marmer et al. 2007; Brown & Yücel 2008) present VAR approaches, where they include weather and storage to test previous findings of the relationship between oil prices and natural gas prices. They find that, when accounting for these factors in

the VAR framework, oil prices maintain their prominent role in explaining natural gas price movements. With these modeling approaches, none of the identification issues are resolved and the parameter estimates are predicated on the idea that all the variables are endogenous to the system. The coefficient estimates represent quantification of the time path of the effects of an unexpected shock of one of the variables of interest on others in the system that is assumed to be moving towards equilibrium. The VAR approach is less helpful in investigating long-term relationships or structural drivers or a permanent change in one of the included variables. For example, the marginal changes are an estimate of the shock working through each variable simultaneously. An exception to this drawback is if any of the variables are deemed to be co-integrated. Cointegration refers to the property between two or more non-stationary variables, where there exists a consistent relationship between the levels of the variables. This property can be tested by regressing one against another and ending up with a stationary error term (Murray 1994). For example, this is the case for the Henry Hub natural gas price and the WTI price for crude oil found in Brown & Yücel (2008). Because of this they were able to estimate a long run relationship between the two.

The Henry Hub price is important for many financial contracts and represents the spot price of natural gas in the most liquid market in the US. The difference between this spot price and the price paid by end users in each region has been the topic of a number of studies. Since the 1970s, the market for natural gas has undergone a number of rounds of deregulation. By 1989, wellhead price ceilings had been

removed. Resultant regional natural gas spot markets emerged and an interstate network of pipelines developed. In 1992, pipelines were required to provide “open-access” transportation, separating the pipeline owners’ ability to sell the commodity and the transportation capacity together (Apergis et al. 2015). De Vany & Walls (1993) argued that a measure of the extent of integration is to test the level of cointegration between wellhead prices and spot markets separated by a pipeline system. They found that in 1987, before the deregulation, cointegration relationships did not exist between most wellheads and spot prices. By 1991, more than 65% had become cointegrated as a result of the deregulation according to the authors. Subsequent studies have used this idea and tested the effect of deregulation and connectedness on the co-integration of a number of price series (Walls 1993; Apergis et al. 2015; Arano & Velikova 2010). They each find that deregulation led to a more integrated market where prices are functionally related to each other in the long run. To what extent pipeline capacity works to increase or decrease the level of integration is not addressed. These papers also do not address a marginal change in capacity on a region’s prices or consumption. Arano & Velikova (2010), for example, ask if citygate (the price the local utilities pay) and residential prices are cointegrated within a state. They find that they are related in the long run post-deregulation supporting the thesis that deregulation has led to better market integration. However, Marmer et al. (2007) use similar cointegration tests to identify constrained regions, whose price time series deviate from long-term relationships with the Henry Hub spot price. Surprisingly, they find that California is the most isolated and that there were no bottlenecks into the Northeast. The constrained conditions have been more recent, while their study’s data

covers 1988-2003. In any case, the methods in the paper would not allow for an estimate of a marginal change in pipeline capacity on price and quantity.

Imposing more structure on a VAR system, typically by limiting directions of causality between variables, allows for estimates of a selection of structural relationships as well as estimates of the transient shock effects estimated from the VAR model. This approach is known as the Structural Vector Auto Regression (SVAR). Nick & Thoenes (2013) present an SVAR model for the German natural gas market. They find that supply side shocks may be overstated when viewed in a model that includes a more comprehensive account for other explanations of price movements, namely abnormal demand scenarios (particularly cold winters). Market fundamentals of storage and weather are found to be significant explanatory variables of the respective natural gas prices that the authors sought to model. While unexpected shocks are important to understand, they tell us little about a permanent change in supply, such as the addition of pipeline capacity to a region.

A recent report prepared for the Department of Energy uses a proprietary modal and agent based model from Deloitte consulting to estimate longer-term (15-year) national and regional prices under various demand scenarios (DOE 2015; Deloitte Center for Energy Solutions 2015). The report supports the need for larger pipeline capacity to constrained regions, such as New England, but does not present marginal impacts of improved capacity or their interaction with other fundamentals explicitly. A more detailed study specific to the New England region was prepared by Competitive

Energy Services (2014), which econometrically modeled daily demand but assumed exogenous prices of the various supply sources (pipeline, LNG, other compressed gas substitutes). The Department of Energy assessed the adequacy of capacity in the Northeast region based on planned pipeline expansions and demand scenarios, but did not consider prices, but rather if the capacity would meet “essential human needs” in terms of quantity in disruption events (DOE 2013). The needs curiously do not include the need for natural gas in electric power generation, but just residential and commercial uses. The study finds that New England and New York City are vulnerable to pipeline disruptions by their definitions.

While including pipeline capacity directly has, to my knowledge, not been included in econometric estimates of New England’s regional natural gas prices or quantities, it is often cited as a major driver in elevated prices for the region in recent winter months (Evans-Brown 2014; Gellerman 2015; National Grid 2014; Ailworth 2014b). There is theoretical and previous empirical justification for this hypothesis using economic theories of congestion. Oliver et al. (2014) present a model of the price differential between two hubs and the congestion of a connecting pipeline to motivate an empirical analysis of a Rocky Mountain region pipeline that closely matches their theoretical model. They find strong empirical support for the effect of congestion on spot prices.

This paper will build on previous work modeling natural gas prices by including relevant fundamental factors of weather, storage, and alternative fuels while adding

and testing the addition of an indicator of pipeline capacity as an important explanatory variable for a regional natural gas market. A description of the data is followed by the econometric methods and results.

Data

The data for the study spans January 2002 to December 2014. Monthly data on natural gas prices and consumption of each state by sector (electric, residential, commercial, industrial) are from the EIA's Natural Gas Monthly reports and obtained through the online data portal. Oil and LNG prices were also obtained through the EIA.

There are multiple prices for natural gas, depending on the market of interest. The city-gate price (Algonquin Citygate for New England) is the wholesale price of gas received by local gas utilities from a pipeline operating company. The sector-specific prices and consumption data comes from surveys reported from individual purchasers and the prices reflect the total cost of the natural gas: the commodity as well as delivery to the end-user (EIA 2015b). These sector prices reflect the type of contractual arrangement that the buyers have with the pipeline or utility for residential, commercial and industrial prices. So, although the Algonquin spot price is a general indicator of the monthly cost of natural gas in the region, it does not reflect the actual price paid by end users.

Pipeline capacity data by state and region is reported annually by the FERC (Federal Energy Regulatory Commission), which is tasked with regulating pipeline use and

construction. FERC also approves future projects and release a projection of capacity based on announced or under construction additions and subtractions. Pipeline capacity is represented in MMcf/d, million cubic feet per day. An estimate of pipeline utilization for a month is the difference between total consumption for the month and the maximum possible based on pipeline capacity. For example, for the year 2014 the northeast states had a net import capacity of 120 bcf/month (billion cubic feet) and a consumption of 98 bcf in January, but only 62 bcf in July. For non-firm contract holders, the capacity remaining after firm demand is met reflects their time-varying supply.

Total monthly consumption by sector is used to separate the firm contract demand (local utilities for residential, commercial and industrial uses) from electric sector gas consumption, which occurs under interruptible contracts. The difference between pipeline capacity and firm demand determines the capacity leftover for the electric sector. This is the capacity variation I use to identify the effect of a marginal increase in capacity on the gas market for the electric energy generators. In essence, this would assume price inelastic short-term demand from firm contract sectors. Exact monthly utilization data reported from each pipeline operator and not gleaned from EIA consumption data would be ideal, but it could not be obtained in a consistent manner and is not made available through public reporting.

Previous studies have found strong relationships between the storage, or deviations from normal storage, and the price of natural gas (Mu 2007; Nick & Thoenes 2013).

Natural gas is stored in underground reserves in depleted oil and gas fields and in places where the right geologic properties for storage naturally occur such as salt caves and aquifers. The amount of gas in storage less the base gas needed to keep the storage site operational is known as the site's working gas reserves. The EIA reports the quantity of gas stored by consuming region. This study uses working gas for the eastern consuming region. The extent to which storage affects regional prices within the context of other constraints will be tested in the SEM econometric specification.

Degree days are defined as the difference between the daily average temperature and 65°F. The monthly measure is the sum of the degree days for the month. Both heating and cooling degree days are expressed as positive values. The idea is that cooling demand is determined by how far above the temperature is from 65°F and the opposite for heating. Degree days and differences from their long-run (1981-2010) averages, or degree day anomalies, are strong explanatory variables for natural gas demand and prices. This data comes from NOAA's National Weather Service Degree Day Statistics database, which provides population-weighted heating and cooling degree days by month, state, and region (NOAA 2015).

Model

The economic system can be summarized by a system of simultaneous equations. There is one equation for each endogenous variable and one identity. Assessing the system in the structural form helps make clear the assumptions of the econometric specifications.

$$Pe_t = \gamma_{10} + \beta_{11}Qe_t + \gamma_{11}CDD_t + \gamma_{12}HDD_t + \gamma_{13}DDd_t + \gamma_{14}Oilp_t + \beta_{12}WG_t \quad (1)$$

$$+ \gamma_{15}t + \varepsilon_{1t}$$

$$Qe_t = \gamma_{20} + \beta_{21}Pe_t + \beta_{22}Ce_t + \gamma_{21}CDD_t + \gamma_{22}HDD_t + \gamma_{23}DDd_t + \beta_{23}WG_t \quad (2)$$

$$+ \gamma_{24}Oilp_t + \gamma_{25}LNGp_t + \gamma_{26}t + \varepsilon_{2t}$$

$$Qo_t = \gamma_{30} + \gamma_{31}CDD_t + \gamma_{32}HDD_t + \gamma_{33}DDd_t + \beta_{31}WG_t + \gamma_{34}Oilp_t \quad (3)$$

$$+ \gamma_{35}LNGp_t + \gamma_{36}t + \varepsilon_{3t}$$

$$WG_t = \gamma_{40} + \beta_{41}Qe_t + \beta_{42}Qo_t + \gamma_{41}CDD_t + \gamma_{42}HDD_t + \gamma_{43}DDd_t + \gamma_{44}t + \varepsilon_{4t} \quad (4)$$

$$Ce_t = Pipes_t - Qo_t \quad (5)$$

The variables are defined as follows:

Table 1

Equation (1) is the natural gas demand equation for the electric energy generating sector and equation (2) is the supply equation. Equation (3) is the firm contract natural gas quantity consumed. Equation (4) represents working gas storage. Equation (5) is the identity making explicit the estimate of capacity left for the electric sector being the difference between firm contract consumption and total pipeline capacity. Including Ce_t , the measure of capacity for the electric energy sector in the econometric specifications, will test its importance in explaining price and quantity in this market.

The simultaneity between prices, quantities and working gas storage makes identifying the structural parameters (γ for exogenous and β for endogenous variables) difficult. For example, to estimate a change in supply as a result of a change in price for natural gas (β_{21}) using OLS, one would have to ensure that the variable for the price (Pe) was uncorrelated with the error term (ε_2) in that equation. Given that the quantity and prices are simultaneous, the correlation is guaranteed. To see this, substitute Qe_t from equation (2) into equation (1) and solve for Pe_t . This algebra involves dividing equation (1), including the error term, through by a term including the parameter β_{11} , after the substitution. Therefore, the error term is correlated with consumption, violating an assumption of OLS leading to bias in the coefficient estimate (Murray 2005). Using OLS to estimate a structural parameter would be a biased representation of the effect of prices on consumption as well as biasing other coefficients in the regression. The same logic holds for the other coefficients on endogenous variables.

The reduced form equations, which represent endogenous variables only as a function of exogenous variables (from each equation), can provide unbiased estimates of the marginal effect of only exogenous regressors on the equilibrium price and quantity of natural gas for the electric sector. The reduced form models can be seen as simplified versions of the structural equation system with substitutions made for each endogenous variable. The path of the effect, such as a shift in demand leading to a change in price, is not identified with reduced form models. However, the reduced form equations are sufficient to estimate the effect of exogenous variables on the equilibrium values of endogenous variable to the system. They estimate changes in

equilibrium values as a result of changes in an exogenous variable since the exogenous variable may directly affect a number of endogenous variables simultaneously, which in turn determine equilibrium values of endogenous variables. By excluding the endogenous variables from each reduced form equation, but including exogenous variables from each structural equation in each estimated reduced form equation, the effect of the exogenous variables are captured, but their path through each endogenous variable is not.

Taking an example of a change in temperature, an exogenous variable to the system, the reduced form equations would estimate the marginal effect on the equilibrium price and quantity for natural gas as a results of simultaneous shifts in both the demand and supply curves due the change in temperature. Whereas, the marginal effects estimated in the SEM model are interpreted as the marginal effect of a variable assuming the other endogenous variables are held fixed. To estimate the corollary equilibrium effects from the SEM model one has to multiply the indirect effects (through changes in other structural equations) and add that to the direct effect from the exogenous variable in the structural equation of interest. An overview and explanation of the use of reduced form models in estimating equilibrium price and quantities is covered in Murray (2005) and a full techincal explanation in Zellner (1974). The primary interest of this study, the effect of a capacity expansion on equilibrium price and quantity for natural gas for the electric sector can be addressed with a reduced from equations of the system. Therefore, I estimated the following reduced form equations:

$$Pe_t = \pi_{10} + \pi_{11}Ce_t + \pi_{12}HDD_t + \pi_{13}CDD_t + \pi_{14}DDd + \pi_{15}Oilp_t + \pi_{16}LNGp_t + \pi_{17}t + \eta_t \quad (6)$$

$$Qe_t = \pi_{20} + \pi_{21}Ce_t + \pi_{22}HDD_t + \pi_{23}CDD_t + \pi_{24}DDd + \pi_{25}Oilp_t + \pi_{26}LNGp_t + \pi_{27}t + \eta_t \quad (7)$$

The capacity left for the electric sector, Ce_t , is exogenous to these two equations assuming Pe_t and Qe_t do not belong in the equation for Qo_t , firm contract demand. Essentially, this assumes firm demand and pipeline capacity are exogenous to the natural gas price and quantity for the electric sector. As the equations are stated in the structural form, there may be an indirect path from Qe_t , to Qo_t through working gas reserves, possibly challenging this assumption. Put another way, if changes in working gas reserves affect both firm and interruptible natural gas price supply and demand, which both in turn affect changes in working gas reserves, the firm demand and leftover capacity may be endogenous to equation (6) and (7) above. We will put this path of effects aside for investigation in the estimation of the SEM model, which can deal with this endogeneity.

In order to obtain unbiased and consistent estimates of the reduced form equations above, a number of additional econometric issues need to be overcome. To meet the assumptions needed of OLS to be unbiased and consistent, the error term must be stationary, not correlated with its lags, and normally distributed. Stationarity refers to

the property of mean, variance, and autocorrelation of a time series that is consistent across time. Price and quantity time series are notoriously non-stationary.

Regressions between non-stationary variables may lead to spurious results if the two variables happen to follow a similar random path across the observed time period (even if the error term of a regression appears stationary). If this is the case, large values of one series do correlate with large values of the others and vice-versa. Running OLS would then assume, incorrectly, that the mean is constant which leads to estimating relationships that are meaningless. Regressions between stationary and non-stationary variables also lead to estimation issues, as the non-stationarity is then passed on to the error term. Therefore, I ran tests for stationarity in the form of an augmented Dickey Fuller test, which tests the presence of a unit root, or a coefficient of 1 on a regression of a variable with the lag of itself. These are presented in Table 2. Those variables where we fail to reject a unit root, implying non-stationarity, are differenced and then re-tested for difference stationarity. The failure to reject a unit root for oil prices and LNG prices implies these are non-stationary series.

The price variables and the measure of leftover capacity are included in natural log form. Therefore, I am estimating a log-log model between those variables and a log-linear relationship between the variables left in levels. Both of these specifications assume a non-linear underlying relationship. Oliver et al. (2014) find this sort of non-linear relationship between the capacity constraint and the price differential between pipeline nodes. How well this type of relationship describes the price and quantity data

in the Northeast will depend on the behavior of the error term in the regression. Namely, after variable transformations, the residuals from the regression should be normally distributed. Using only non-logged data led to a non-normally distributed error term and a fitted model that underestimated the price increases in constrained months. The quantile plots of the residuals of both specifications are found in Figure 4 showing the residuals to be closer to normally distributed (the straight lines) using the log transformations.

In each specification of the model using OLS, there exists month-to-month autocorrelation in the residual, which would bias the estimates of the coefficient standard errors. Although the inclusion of the degree-day variables captures much of the seasonality in the relationships, there exists a remaining seasonal correlation component of the error term. To overcome these two autocorrelation issues, I include a seasonal lag of the dependent variable and handle the remaining autocorrelation process using GLS (generalized least squares) using maximum likelihood estimation methods. First, I present the OLS estimates with no consideration for non-stationary variables followed by including them in differenced form. The preferred models, one for price and one for quantity, are the furthest right in Tables 3 and 5.

GLS is well suited to handling time series data by modeling autocorrelation in the residual in order to obtain unbiased estimates of the standard errors of the coefficients for statistical inference of marginal effects. ARIMA (autoregressive integrated moving average) are closely related and are often used for near-term forecasting of economic

time series. The method relies on past history of a time series to make prediction of future values. Without the inclusion of exogenous regressors (ARIMAX in the parlance of these models), the method is considered atheoretical, as there exists no underlying theory as to why the series should behave in the way described (Hyndman & Athanasopoulos 2013). With the inclusion of exogenous regressors, the models become nearly equivalent to a GLS specification if the error lag structure is specified similarly. In the end, both are estimated using distributional assumptions and maximization of a log likelihood function to fit the parameters. Therefore as a robustness check, I performed an ARIMA analysis. The marginal effect results are similar, but the simulation predictions are slightly different due to how the error term is treated in step ahead forecasting. The model and results from this analysis are in the Appendix.

Reduce Form Results

Table 3 presents the results of the reduced form price regressions. Columns 1-4 show the OLS results with additions of oil prices, the seasonal lag, and LNG prices. Column 5-6 shows the GLS estimates taking into account the remaining autocorrelation in the OLS error term. I specified an ARMA (1,1) process based on comparing the AIC from alternative model specifications for the residual autocorrelation. The resulting GLS estimates show a well-behaved, stationary error term with no autocorrelation. The same process was repeated for column 6, where the non-stationary regressors were replaced with the differenced series.

The results show a significant relationship between the pipeline capacity left for the electric sector after taking into account firm demand and the price of gas for this sector. The coefficient (-.48) means that an increase in capacity left of 1% lead to a .48% decrease in price. The coefficient on degree day anomalies is significant and positive. Degree day deviations remain in level form, so a 1 degree change in degree day anomalies for a month leads to .06% increase in price.

The coefficients on heating and cooling degree days are insignificant. These are highly correlated with the measure of leftover capacity, which is itself a function of degree days through firm demand, equation (3). Including both raises issues of collinearity. The coefficient correlation matrix reveals that the estimates of the coefficients of HDD, and CDD are in fact correlated with the leftover capacity, our variable of interest (correlations of .82 and .24 respectively). Specifications with combinations of including temperature and capacity variables together and separately are in Table 4 as robustness checks. One would expect collinearity to inflate standard errors of coefficient on both variables, inflate t-values and thus I would fail to reject that coefficients are different then zero more often. The significance of the coefficient on leftover capacity in the model including both leftover capacity and temperature measures remains strong. This means that it is to a larger extent the congestion that drives the price electric generators pay. It also supports modeling the leftover capacity once the firm demand is accounted for when estimating natural gas prices for the electric sector.

That degree day anomalies were significant and not degree days itself may be partly a result of including a seasonal lag, which takes up much of the variation in monthly average temperatures. It is the deviations from normal conditions that provide explanation above and beyond the average monthly conditions. I left the working gas reserve variable out of the regression due to the simultaneity issues. Thus, the estimates of the effect of the temperature variables could be working through changes in working gas as well as through other excluded endogenous variables in the reduced form estimates. In order to separate those effects, one would have to estimate the whole system of equations with identifying restrictions.

Table 5 shows the results for the reduced form equilibrium quantity estimates. The coefficients are of the expected sign. The seasonal pattern of natural gas consumption for electricity generation means more is consumed in the summer when prices are low and demand is high and less is consumed in the winter with constrained capacity and high prices (Figure 1). The positive and significant coefficient on CDD reflects the summer cooling demand. Since the equation is in reduced form, the negative sign on degree day anomalies may reflect the effect of the resultant increase in price that go along with those conditions. Again, a system of equations with identifying restrictions would be needed to disentangle these two effects. Lastly, our variable of interest, leftover capacity is significant and positively related to the amount consumed by electric energy generation.

Increase in Capacity Scenario Analysis

The estimates of equilibrium price and quantity allows for estimating the effect of an increase in pipeline capacity to the region. Holding other covariates constant, a 1% increase in pipeline capacity would lead to .48% decrease in price and an increase in consumption of .21%. Price and quantity series for a number of weather scenarios are estimated as follows. The scenarios are assessed under current and future pipeline capacities. To do so, first I estimate the leftover capacity as a function of the quantity of gas consumed by firm contract holders, equation (5) in the system of equations above. Again, I am assuming that the price and quantity consumed by the electric energy sector is exogenous to the firm quantity demand Qo_t . As a result, I can estimate this separately, as neither Pe_t , or Qe_t appear in the equation for leftover capacity, Ce_t .

The regression results for estimating Qo_t are in Table 6. I use the same process as the previous regressions. First, I fit the OLS model with the covariates, determine the error structure of the residuals, and then re-estimate with GLS accounting for the error structure (ARMA(2,1)). The model is a good fit with a small standard error (.12), which is a 1.1% percent error. The difference between this estimate of firm demand and the pipeline capacity, equation (5), creates the capacity left for the electric sector.

Based on current proposed expansion on the Algonquin and Tennessee pipelines, a marginal increase in capacity of 450 MMcf/d, about an 11% increase in total capacity into New England is currently planned (ICF International 2014). The expansion on the Algonquin line is largely along existing routs and is comprised of increasing the

diameter of the pipeline in certain sections as well as upgrading compressors stations and adding a few auxiliary loops (FERC 2015). The Tennessee pipeline expansion involves siting new routes from New York through western Massachusetts to the greater Boston, MA area (Kinder Morgan 2015). Both projects are making their way through the review and regulatory process and face opposition for their impacts on the planned routes and the environmental implications of the fossil fuel energy (Ailworth 2014a; Jay Fitzgerald 2014; Zeller 2014; Marc Levy 2014). They are both slated to be completed in 2016. While the timing and exact capacity of these particular project will likely change over time, this study will use the 450 MMcf/d as the increase in pipeline capacity for the scenarios and to estimate the effect under scenarios of winter weather.

To generate the weather scenarios, temperature series for a selection of years are made from the temperature dataset. The degree day data from NOAA in this form goes back to 1997. On the basis of cumulative HDD, the coldest winter was that of 2003/4, followed closely by the winter of 2013/14. The winter of 2003/04 is used as the severe cold scenario, 2008/09 the median and 2011/12 the low demand scenario. I simulate a year window, for example from September 2003 to the end of August 2004, to capture the path of the equilibrium price and quantity through a continuous winter. By presenting a range of winter severities, the impact of the expansions is shown under a range of possible weather conditions. The other exogenous covariates, oil prices and LNG prices, neither of which exhibit seasonality, are held fixed at their averages, and thus differenced to equal zero, for the scenarios. The lagged dependent variables, natural gas prices for electric power generation and total consumption from all other

sectors are the 2014 price series.

Figure 5 and Tables 7 and 8 show the price and quantity projections for the winter scenarios under the current and expanded pipeline capacities. The figures show price in solid lines and quantity in dashed lines for both the current (black) and future (red) capacity scenarios. Prices are the highest in January when the system is constrained by high heating demand from firm contract holders, while the quantity consumed is highest in the summer months when price is low and electric energy demand is high. The severity of the winter increases the price spike as well as increasing the difference between the two capacity scenarios. For the mild winter, there is a difference in price of \$1.16/Mcf in January, while under the severe winter scenario this increases to \$2.96/Mcf. The quantity projections are inversely related to the price. The increase in capacity alleviates the congestion but is relatively more important in colder winters. In terms of the difference in total expenditure (price times quantity) of natural gas consumed by the electric sector, the increased capacity reduces overall total expenditure in the electric sector by just over \$35 million for the month of January alone. This reflects the lower price and slightly increased quantity consumed. The difference is \$31 million under the median January and \$15 million under the mild scenario.

SEM Model

The reduced form models are useful in estimating equilibrium price and quantity, but have their drawbacks. I was unable to include working gas reserves, a possibly important explanatory variable, directly due to endogeneity. The reduced form coefficients are the total, direct plus indirect, effects between the variables in the system. Thus, I may be attributing an effect to degree days, for instance, that is the sum of degree day's direct effect on demand as well as its indirect effect through changes in working gas reserves, if working gas reserves effect supply or demand directly. Additionally, the structural parameters describe the shape of the demand and supply curve. Of particular interest are the price elasticities. Disentangling these effects by estimating the structural parameters that describe the supply and demand curve for natural gas in this market requires estimating the SEM model.

Simultaneous equation models are used to identify the path of effects of a change in an exogenous variable on endogenous variables in the system. They are used to identify the direct and indirect effects (through other endogenous variables) in order to infer causality from one variable to another. For example, if the effect of increased capacity shifted both the supply of natural gas, but also the storage, which may lead to a change in equilibrium price, an SEM model could quantify what portion of the effect of the change in capacity on prices was due to a change in storage versus a change in supply. The reduced form model estimates only the total effect, thus I could not separate the direct and indirect (through storage) effects. Therefore, for causal inference, which relies on estimating the effect on one variable conditional on all other held constant, SEMs are superior. A reduced form model does not assume all other variables are held

constant and thus the estimated marginal effects are exposed to issues of indirect causality explained above. This is important for estimating the shape of the demand or supply curve, but not necessary for estimating changes in equilibrium values. Equilibrium values are, by definition, the outcome of all the endogenous variables changing simultaneously. Therefore, to identify the price elasticities of supply and demand (shape of the curves), I use an SEM model. To estimate changes in equilibrium values as a results of an expansion of capacity, I use a reduced form model. Additionally, I compare the total effect from the SEM model to the reduced form. Analytically these are the same, but the methods to estimate the equations are slightly different.

I estimate the dynamic simultaneous equations model (SEM) of the system presented earlier. I include the one-month lag of each endogenous variable as well as one seasonal lag term to its respective equation to handle the time series nature of the data. This was based on autocorrelation function plots of the residuals post estimation. After the lags were added, there are no significant lags in the errors term from each equation.

To identify the structural parameters, there needs to be included at least as many exogenous variables as there are equations for endogenous variables in the system. I use the temperature series, oil and LNG prices as well as the lags of each endogenous variable as instruments in the 3SLS estimation of the system. Therefore, my model is over-identified with 13 exogenous or predetermined variables, including the lags, to 4

structural equations. The lagged endogenous variables are considered predetermined, therefore not correlated with the current period's error term and exogenous to the current period (Fox 2002; Henningsen et al. 2007). In addition, a minimum of one exogenous variable must be left out of each equation for identification. This is the SEM version of the exclusion restriction in instrumental variable estimates.¹²

The 3SLS results are presented in Table 9. Each column presents the results of the structural equations described in the model. 3SLS combines 2SLS, the system of equations version of instrumental variables, and SUR (seemingly unrelated regression), which takes into account the remaining correlation between the error terms across each equation and is estimated with generalized method of moments (GMM). 3SLS is therefore more efficient than 2SLS. The null hypothesis of the Hausman test is that the exogenous variables are uncorrelated with the errors term from each equation. Under the null hypothesis both 2SLS and 3SLS are consistent, while under the alternative hypothesis 2SLS is consistent while 3SLS is not. I am unable to conduct a Hausman test due to computational issues in the R package `systemfit`. Workarounds are currently being explored. In the meantime as a robustness check, the 2SLS results are shown in Table 10. The 2SLS are less efficient but unbiased in the case where the exogenous variables are correlated with the error terms. The results are very close, showing similar magnitudes and significance in important relationships although the price elasticity of demand is only significantly different than zero at the 10% level.

¹² Formally, The order condition (necessary for identification), as many exogenous variables as endogenous variables in the system, is therefore satisfied. I also checked that the rank condition (sufficient for identification) is also satisfied given the system specified.

The coefficients for heating degree days and degree day anomalies are positive and significant in the demand equation, reflecting demand being driven by temperature. The variable of interest, capacity for the electric sector, is significant in the supply equation. An increase of capacity of one percent leads to an increase of .7% in quantity supplied, capturing the shift in the supply curve. While the reduced form regressions quantified the total effect on equilibrium price and quantity, the SEM model made explicit the path of the effect, through shifts in the supply curve. The price elasticity of demand is $-.57$ and significant. Thus the total effect of a 1% change of capacity on equilibrium price, the multiplication of the indirect effects of capacity on supply and quantity supplied on demand, is $-.404\%$. This should be, and it is in this case, close to the reduced form estimate. The advantage of the SEM is now I can identify the path of the effect and the shape of the supply and demand curve. The supply price elasticity is small, negative and significant, which is against what theory would imply. However, very low and sometimes negative supply price elasticities for natural gas, especially over the short run, have been found before, but are not common (Arora 2014). The elasticity results show that both supply and demand are price inelastic over this period. The inelastic demand and supply price elasticities means that small excess supplies or demands are only cleared with a large change in price, which helps explain the volatility of the price for natural gas, especially for interruptible contract holders.

Previous work has found changes in natural gas storage conditions to be significant in describing price movements in reduced form models (Brown & Yücel 2008; Mu 2007), but by including them, each assumes that their measure of storage levels is exogenous to the current price of natural gas. If this is not the case, the coefficients from those regression results are biased. The inclusion of working gas reserves to the SEM as an endogenous variable led to unexpected results. Changes in working gas reserves was insignificant in explaining shifts in the supply curve but positively and significantly related to demand. The results of the working gas structural equation are also unexpected. Changes in reserves are positively related to electric energy consumption of natural gas, which would naturally be a draw on reserves. Working gas reserves are also, curiously, significantly positively related with both HDD and CDD. The effect of degree day anomalies are negative and significant which may reflect the difference from average storage conditions in times of particularly high demand. The results raise more questions than answers for the effect of natural gas storage in the market for the electric generating sector.

The way the variation in working gas explains the dynamics of this system appears to be against what theory would imply. One would expect larger working gas reserves to decrease price by increasing available supply. While I included demands and therefore withdrawals on the reserves, explanatory variables for the variation in additions to working gas storage from producers were not included and could be a direction of future inquiry. Since the seasonal pattern is handled by the seasonal lagged term and the temperature variables, variation in additions to storage above and beyond normal

conditions is assumed to be random noise. The gas is being stored to meet expectations of future demand, so if explanatory variables of storage additions are correlated with already included variables to the system, the results could be biased.

Between the curious supply price elasticities and the results of the working gas variable, I do not put much weight on the SEM results at this point other than to reinforce that leftover capacity is a possibly significant driver of supply for this natural gas consuming sector and that the system appears to be price inelastic. These low price elasticities are understandable as supply is fixed with pipeline capacity and demand the result of running large fixed capital.

Discussion

The results of the reduced form regressions show that capacity constraints are a driver of the increase in the price of natural gas in cold winter months. The SEM results support the case for including the capacity constraints as an explanatory variable of the supply of gas to a constrained region such as New England.

Capacity additions represent a shift in the supply curve for natural gas, as identified through the SEM results. The total effect (direct and indirect) of a 1% increase in capacity of prices from the SEM results was similar, but slightly smaller, to that found with the reduced form model, -.44 and -48% respectively. Analytically, these marginal effects on equilibrium prices should be identical assuming the structural model is the correct model of the system. The reduced form model is just a simplification of the

structural model. However, the reduced form models are estimated with GLS using maximum likelihood methods, while the 3SLS estimate with GMM. While both are consistent, meaning they converge to the true value of a parameter they do not have identical small sample properties. Additionally, the SEM model explicitly includes the working gas reserves, which could provide a back-door route for electric sector demand to effect firm demand and thus challenge an assumption needed for the reduced form models to be unbiased. This route, working gas reserves effecting firm demand, is insignificant in the SEM equations. The two approaches produced similar equilibrium results, but the SEM model allowed for the identification of the shape of the supply and demand curves.

To a large extent, the entire system is driven by temperature dependent demand, so identifying the effect of the capacity constraint above and beyond the effect of the temperature variation was always going to be a tall task. That the variable proposed in this study to represent the leftover capacity for the non-firm contract holders remained significant through many specifications, including ones with other correlated covariates, supports the inclusion of this fundamental factor.

The inclusion of the capacity measure and its significance to explaining both price and quantity provides a path to estimate the effect of increasing the capacity of pipelines into New England. This study considers both equilibrium price and quantity changes based on the marginal effects of constrained conditions through the study period. Other approaches have estimated the effect of pipeline additions for New England

(Competitive Energy Services 2014; DOE 2013; ICF International 2014) which either assume price schedules or do not allow for feedbacks between price and quantity, or only consider price not quantity or vice versa. Studies up until this point have left an incomplete picture for the consequences for the electric energy sector. These results quantify the effect when both supply and demand are taken into account to come to an estimate of the change in total expenditure on natural gas by the electric energy generators.

Temperature is a significant driver of price and quantity variations, which is in line with studies of other natural gas markets (Mu 2007; Stephen P. A. Brown & Mike K. Yucel 2008; Nick & Thoenes 2013). While quantities consumed reflect fundamental demand factors, prices reflect scarcity and the extent to which conditions surprise market participants. Large degree day anomalies capture conditions that are relatively more rare and may misalign with market participants expectations of a month's conditions and drive up or down prices to a larger extent. I estimated the effect of increased capacity under a range of winter severities. January, in particular, over the study period led to constrained conditions and large price increases. Depending on the severity of the winter month, the effect of the capacity additions of 450 MMcf/d in terms of total expenditure by the electric sector varies by over 100% (15 Million to 31 Million) for a single month. The non-linear relationship between temperature conditions and the benefits of additional capacity makes using average conditions misleading. Considering the benefits under more severe winters, or with a

probabilistically weighted expected benefit may provide directions moving forward to estimate the benefits of additional pipeline capacity.

Given this non-linearity, pipeline capacity affects the range of prices, reducing the winter spikes to a larger extent in abnormal conditions. Thus, pipeline capacity reduces the price risk of energy supply for the region. However, excess capacity to address winter constrained conditions will come as an additional cost in other times, where the large fixed cost of the system are spread out over the smaller gas volumes that are flowing. A weakness of the methods used in this study is that they do not address directly the possible price changes due to the effect of larger fixed costs of pipeline operation. To what extent the firm demand or interruptible customers share the burden of the increased fixed cost is unclear.

The objective of this paper was to estimate changes in equilibrium prices and quantities of natural gas consumed by the electric sector as the results of a change in pipeline capacity. It is the equilibrium prices for natural gas that describe the per-unit cost of the input to electric generation. These costs are passed along to electric ratepayers. Thus, to estimate the effect of increased capacity on ratepayers, and inform the debate about how the cost of electric energy might change, a reduced form model is well suited. Translating the price for gas that natural gas generators pay to the cost of electric energy to residents and businesses is left for future work. The SEM results clear up the causal mechanism, a shift in supply, and confirm a similar total effect as

from the reduced form models. In addition, the SEM model shows that both the supply and demand curves are price inelastic.

This study provides one piece of the puzzle in determining the effects of proposed expansions, but not the whole picture. In the debate about the impact of proposed pipelines, this study offers a way to quantify the effect on the electric energy generators input price and quantity for natural gas. The benefits to the electric energy sector are potentially large by allowing larger quantities of gas to be consumed at lower price for generation. These benefits should be weighed against the costs of the planned routes, the environmental consequences, natural gas price in other sectors and other indirect benefits and costs.

REFERENCES

- Ailworth, 2013. Wholesale Gas Prices Rise in N.E. *Boston Globe*. Available at: <http://www.bostonglobe.com/business/2013/08/01/tight-pipeline-capacity-pushes-natural-gas-prices-higher/vApWO3InNMvmgeHx93QZ7O/story.html#comments> [Accessed December 5, 2014].
- Ailworth, E., 2014a. Company proposes natural gas pipeline to cross Massachusetts - The Boston Globe. *BostonGlobe.com*. Available at: <https://www.bostonglobe.com/business/2014/06/08/company-proposes-natural-gas-pipeline-cross-massachusetts/uvoxHT3zECRockdVzTIADO/story.html> [Accessed April 28, 2015].
- Ailworth, E., 2014b. Natural gas prices spike - The Boston Globe. *BostonGlobe.com*. Available at: <https://www.bostonglobe.com/business/2014/01/10/natural-gas-prices-spike/H8d7cJyAbnE8cnNliWzEHN/story.html> [Accessed April 28, 2015].
- Apergis, N., Bowden, N. & Payne, J.E., 2015. Downstream integration of natural gas prices across U.S. states: Evidence from deregulation regime shifts. *Energy Economics*, 49, pp.82–92.
- Arano, K. & Velikova, M., 2010. Estimating the long-run equilibrium relationship: The case of city-gate and residential natural gas prices. *Energy Economics*, 32(4), pp.901–907.
- Arora, V., 2014. Estimates of the price elasticities of natural gas supply and demand in the United States. Available at: <http://mpira.ub.uni-muenchen.de/54232/> [Accessed May 28, 2015].
- Brown, S.P.A. & Yücel, M.K., 2008. Deliverability and regional pricing in U.S. natural gas markets. *Energy Economics*, 30(5), pp.2441–2453.
- Competitive Energy Services, 2014. Assessing Natural Gas Supply Options for New England and Their Impacts of Natural Gas and Electricity Prices.
- Deloitte Center for Energy Solutions, 2015. Deloitte Market Point. Available at: http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_er_marketpoint_marketbuilder011411.PDF.
- DOE, 2013. Assessment of the Adequacy of Natural Gas Pipeline Capacity in the Northeast United States.

- DOE, 2014. July 28 Stakeholder Meeting on Natural Gas – Electricity Interdependence. Available at: http://energy.gov/sites/prod/files/2014/07/f17/qermeeting_denver_background_memo.pdf.
- DOE, 2015. Natural Gas Infrastructure Implications of Increased Demand From the Electric Power Sector.
- EIA, 2015a. About U.S. Natural Gas Pipelines. Available at: http://www.eia.gov/pub/oil_gas/natural_gas/analysis_publications/ngpipeline/regulatory.html [Accessed April 1, 2015].
- EIA, 2014. New England relying more on natural gas along with hydroelectric imports from Canada. Available at: <http://www.eia.gov/todayinenergy/detail.cfm?id=17671> [Accessed December 9, 2014].
- EIA, 2013. Northeast grows increasingly reliant on natural gas for power generation. *Today in Energy*. Available at: <http://www.eia.gov/todayinenergy/detail.cfm?id=13751> [Accessed December 4, 2014].
- EIA, 2015b. Table Definitions, Sources, and Explanatory Notes. Available at: http://www.eia.gov/dnav/ng/TblDefs/ng_pri_sum_tbldef2.asp [Accessed April 10, 2015].
- Evans-Brown, S., 2014. New England Electricity Prices Spike As Gas Pipelines Lag. *NPR.org*. Available at: <http://www.npr.org/2014/11/05/361420484/new-england-electricity-prices-spike-as-gas-pipelines-lag> [Accessed April 28, 2015].
- FERC, 2015. Algonquin Incremental Market Project: Final Environmental Impact Statement.
- Fox, J., 2002. Structural equation models. *appendix to an r and s-plus companion to applied regression, McMaster University, Canada*, <http://cran.r-project.org/doc/contrib/Fox-Companion/appendixsems.pdf>. Available at: <http://www.vps.fmvz.usp.br/CRAN/doc/contrib/Fox-Companion/appendix-sems.pdf> [Accessed May 28, 2015].
- Gellerman, B., 2015. As Winter Electricity Prices Jump, Mass. Debate Over Natural Gas Pipelines Heats Up. *wbur*. Available at: <http://www.wbur.org/2015/03/10/natural-gas-pipelines> [Accessed April 28, 2015].

- Henningsen, A., Hamann, J.D. & others, 2007. systemfit: A package for estimating systems of simultaneous equations in R. *Journal of Statistical Software*, 23(4), pp.1–40.
- Hyndman, R.J. & Athanasopoulos, G., 2013. *Forecasting: principles and practice*, S.I.: OTexts.
- ICF International, 2014. Assessment of New England’s Natural Gas Pipeline Capacity to Satisfy Short and Near-Term Electric Generation Needs: Phase II.
- ISO New England, 2015. ISO on Background: State of the Grid - Managing a System in Transition.
- Jay Fitzgerald, 2014. In face of opposition, company to reroute pipeline. *The Boston Globe*. Available at: <http://www.bostonglobe.com/business/2014/12/05/face-opposition-company-reroute-pipeline/wj0k4WbfYr5FFyyHtPmFGJ/story.html> [Accessed December 8, 2014].
- Kinder Morgan, 2015. Kinder Morgan Northeast Expansion Project Open Season. Available at: http://www.kindermorgan.com/business/gas_pipelines/east/neenergydirect/ [Accessed April 20, 2015].
- Marc Levy, 2014. U.S. Shale Gas Line to the Northeast Gets Federal Nod. *Associated Press*. Available at: <http://abcnews.go.com/US/wireStory/shale-gas-pipeline-northeast-fed-nod-27343406> [Accessed December 8, 2014].
- Marmer, V., Shapiro, D. & MacAvoy, P., 2007. Bottlenecks in regional markets for natural gas transmission services. *Energy Economics*, 29(1), pp.37–45.
- Mohammed A. Al-Sahlawi, 1989. The Demand for Natural Gas: A Survey of Price and Income Elasticities. *The Energy Journal*, 10(1), pp.77–90.
- Morris, G., 2013. Long-Term Contracts Remain Elusive In Power Sector. Available at: <http://www.aogr.com/web-exclusives/exclusive-story/long-term-contracts-remain-elusive-in-power-sector> [Accessed April 1, 2015].
- Murray, M.P., 1994. A Drunk and Her Dog: An Illustration of Cointegration and Error Correction. *The American Statistician*, 48(1), p.37.
- Murray, M.P., 2005. *Econometrics: A Modern Introduction* 1 edition., Boston: Prentice Hall.
- Mu, X., 2007. Weather, storage, and natural gas price dynamics: Fundamentals and volatility. *Energy Economics*, 29(1), pp.46–63.

- National Grid, 2014. National Grid Files for Winter Rates Massachusetts. Available at: http://www.nationalgridus.com/aboutus/a3-1_news2.asp?document=8764 [Accessed December 4, 2014].
- Nick, S. & Thoenes, S., 2013. *What drives natural gas prices? A structural VAR approach*, EWI Working Paper. Available at: <http://www.econstor.eu/handle/10419/74410> [Accessed December 3, 2014].
- NOAA, 2015. Climate Prediction Center - Monitoring & Data: Degree Days Statistics. Available at: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/cdus/degree_days/ [Accessed April 10, 2015].
- Oliver, M., Finnoff, D. & Mason, C., 2014. Optimal Capacity and Two-Part Pricing for Natural Gas Pipelines under Alternative Regulatory Constraints. Available at: http://www.usaee.org/usaee2014/submissions/OnlineProceedings/Oliver_Finnoff_Mason_Feb2014.pdf [Accessed May 13, 2015].
- Oliver, M., Mason, C. & Finnoff, D., 2014. Pipeline Congestion and Natural Gas Basis Differentials: Theory and Evidence. *Journal of Regulatory Economics*, 14(3), pp.261–291.
- Pfaff, B., 2008. VAR, SVAR and SVEC models: Implementation within R package vars. *Journal of Statistical Software*, 27(4), pp.1–32.
- Stephen P. A. Brown & Mike K. Yucel, 2008. What Drives Natural Gas Prices. *The Energy Journal*, 29(2), pp.45–60.
- De Vany, A. & Walls, W.D., 1993. Pipeline Access and Market Integration in the Natural Gas Industry: Evidence from Cointegration Tests. *The Energy Journal*, 14(4), pp.1–19.
- Walls, W.D., 1993. Competition, Prices, and Efficiency in the Deregulated Gas Pipeline Network: A Multivariate Cointegration Analysis. *Journal of Energy and Development*, 19(1). Available at: <http://www.osti.gov/scitech/biblio/229903> [Accessed May 13, 2015].
- Zeller, T., 2014. Natural Gas Pipeline Plan Creates Rift in Massachusetts. *The New York Times*. Available at: <http://www.nytimes.com/2014/07/11/business/energy-environment/new-england-confronts-surgin-demand-for-natural-gas.html> [Accessed April 28, 2015].
- Zellner, A. & Palm, F., 1974. Time Series Analysis and Simultaneous Equation Econometric Models. *Journal of Econometrics*, 2(1), pp.17–54.

Tables and Figures

TABLE 1
Variable Definitions

Symbol	Name	Units	Exogenous
Pe	NG price for electric sector	\$	
Qe	NG quantity for electric sector	MMcf	
Ce	NG capacity for electric sector	MMcf	
Qo	NG consumed by other sectors	MMcf	
WG	Working gas reserves	MMcf	
HDD	Heating degree days	°F	x
CDD	Cooling degree days	°F	x
DDd	Degree day anomaly	°F	x
$Oilp$	WTI oil price	\$	x
$LNGp$	LNG import price	\$	x
$Pipes$	New England pipeline capacity	MMcf	x

Note: Price, consumption and capacity data are from the EIA's Natural Gas Monthly. Degree day data are from NOAA's National Weather Service Degree Day Statistics database. The last column indicates the assumptions of exogeneity of the variables to the system for identification.

TABLE 2
Unit Root Tests

Augmented Dickey-Fuller		
	<i>Levels</i>	<i>Differences</i>
NG price for electric	-3.6283**	
Leftover capacity	-5.49***	
Degree days	-17.61***	
Degree day diff	-5.09***	
Oil price	-2.36	-5.9732***
Working gas diff	-3.33*	-5.1506***
Working gas	-7.57***	
LNG price	-2.53	-5.6432***
* p<.10 ** p<.05 *** p<0.01		

Note: The test statistic for the Augmented Dickey-Fuller test of a unit root is presented for both the level data as well as the differenced data. The null hypothesis of a unit root, implying non-stationary is rejected with significant test statistics. Variables are in levels or first differences.

TABLE 3
Equilibrium Natural Gas Price
New England Electric Sector

<i>Dependent variable:</i> Log NG Price Electric						
	<i>OLS</i>				<i>GLS</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
Log NG Capacity	0.030 (0.201)	-0.586** (0.235)	-0.979*** (0.214)	-0.595*** (0.176)	-0.563*** (0.187)	-0.479** (0.186)
HDD	0.00048*** (0.0002)	0.00010 (0.0002)	-0.00026 (0.0002)	-0.00004 (0.0001)	0.00001 (0.0001)	0.00005 (0.0001)
CDD	0.00058 (0.001)	0.00013 (0.001)	0.00003 (0.001)	0.00023 (0.0004)	0.00021 (0.0003)	0.00028 (0.0003)
DDd	0.00121*** (0.0004)	0.00128*** (0.0004)	0.00103*** (0.0004)	0.00066** (0.0003)	0.00060*** (0.0002)	0.00058*** (0.0002)
Log Oil Price		0.356*** (0.081)	0.191** (0.080)	0.017 (0.067)	0.171 (0.139)	
Seasonal Lag			0.406*** (0.070)	0.239*** (0.059)	0.301*** (0.076)	0.280*** (0.080)
Log LNG Price				0.629*** (0.071)	0.123** (0.061)	
Δ Log Oil Price						0.060 (0.163)
Δ LNG Price						0.003 (0.004)
Observations	156	156	144	144	144	144
R ²	0.209	0.300	0.427	0.638		
Adjusted R ²	0.188	0.277	0.401	0.620		
Log Likelihood					46.593	45.653
Akaike Inf. Crit.					-71.185	-69.306
Bayesian Inf. Crit.					-38.594	-36.638
Residual Std. Error	0.386 (df = 151) 9.985***	0.364 (df = 150) 12.883***	0.318 (df = 137) 16.988***	0.255 (df = 135) 34.049***		
F Statistic	(df = 4; 151)	(df = 5; 150)	(df = 6; 137)	(df = 7; 135)		
* p<.10 ** p<.05 *** p<0.01						

Note: Each column represents a different specification of the model explaining natural gas prices paid by the electric energy sector. The last 2 columns are from GLS regressions with ARMA(1,1) errors specified based on the residuals of the OLS regression. Observation lengths differ to the inclusion of a seasonal lag. Oil and LNG prices are in first differences in the last column due to non-stationarity issues.

TABLE 4
Test of Treatment of Temperature Variables
Equilibrium Natural Gas Price for New England Electric Sector

<i>Dependent variable:</i>					
Log NG Price Electric					
	(1)	(2)	(3)	(4)	(5)
Log NG Capacity			-0.47868** (0.18643)	-0.56018*** (0.10376)	-0.50246*** (0.18817)
HDD		0.00038*** (0.00009)	0.00005 (0.00014)		0.00008 (0.00014)
CDD		0.00051* (0.00029)	0.00028 (0.00029)		0.00024 (0.00035)
DDd	0.00079*** (0.00016)	0.00048*** (0.00018)	0.00058*** (0.00019)	0.00065*** (0.00017)	
HDDd					0.00055*** (0.00019)
CDDd					0.00091 (0.00063)
Seasonal Lag	0.03009*** (0.00913)	0.02609*** (0.00898)		0.02876*** (0.00884)	0.02826*** (0.00894)
Δ Log Oil Price			0.28047*** (0.07951)		
Δ LNG Price	0.10587 (0.15612)	0.04734 (0.16058)	0.05992 (0.16323)	0.06945 (0.16083)	0.01851 (0.16679)
Observations	144	144	144	144	144
Log Likelihood	13.718	6.764	9.250	23.793	2.970
Akaike Inf. Crit.	-11.436	6.473	3.501	-29.586	18.060
Bayesian Inf. Crit.	12.040	35.673	35.540	-3.241	52.923
*p<.10 **p<.05 ***p<0.01					

Note: GLS regressions of variations of the preferred natural gas price model are presented. Column 3 is the preferred model, column 6 in Table 3. The measure of capacity for the electric energy sector remains significant through alternative specifications of the model with combinations of correlated temperature variables

TABLE 5
Equilibrium Natural Gas Quantity
New England Electric Sector

<i>Dependent variable:</i>						
Log NG Quantity for Electric						
	<i>OLS</i>				<i>GLS</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
Log NG Capacity	0.522*** (0.072)	0.443*** (0.094)	0.293*** (0.079)	0.266*** (0.078)	0.268*** (0.097)	0.214** (0.090)
HDD	0.00019*** (0.0001)	0.00014* (0.0001)	0.00009 (0.0001)	0.00008 (0.0001)	0.00006 (0.0001)	0.00003 (0.0001)
CDD	0.00178*** (0.0002)	0.00172*** (0.0002)	0.00143*** (0.0002)	0.00141*** (0.0002)	0.00139*** (0.0002)	0.00143*** (0.0002)
DDd	-0.00016 (0.0002)	-0.00017 (0.0002)	-0.00028** (0.0001)	-0.00021 (0.0001)	-0.00018 (0.0001)	-0.00018 (0.0001)
Log Oil Price		0.040 (0.030)	-0.057* (0.030)	-0.021 (0.032)	-0.045 (0.046)	
Seasonal Lag			0.293*** (0.064)	0.286*** (0.063)	0.211*** (0.065)	0.193*** (0.062)
Log LNG Price				-0.085*** (0.030)	-0.011 (0.036)	
Δ Log Oil Price						-0.165 (0.102)
Δ LNG Price						0.005 (0.003)
Observations	168	168	156	155	155	156
R ²	0.607	0.611	0.695	0.710		
Adjusted R ²	0.597	0.599	0.683	0.696		
Log Likelihood					135.010	137.551
Akaike Inf. Crit.					-246.020	-251.101
Bayesian Inf. Crit.					-209.499	-214.503
Residual Std. Error	0.151 (df = 163)	0.151 (df = 162)	0.122 (df = 149)	0.119 (df = 147)		
F Statistic	62.860*** (df = 4; 163)	50.868*** (df = 5; 162)	56.584*** (df = 6; 149)	51.371*** (df = 7; 147)		
* p<.10 ** p<.05 *** p<0.01						

Note: Each column represents a different specification of the model explaining natural gas quantity consumed by the electric sector. The last 2 columns are from GLS regressions with ARMA(2,1) errors specified based on the residuals of the OLS regression. Observation lengths differ due to the inclusion of a seasonal lag. Oil and LNG prices are in first differences in the last column due to non-stationarity issues.

TABLE 6
Estimate of Natural Gas Quantity
New England Firm Contract Quantity

	<i>Dependent variable:</i>		
	Firm Contract Quantity Consumed		
	<i>OLS</i>		<i>GLS</i>
	(1)	(2)	(3)
HDD	0.00120*** (0.00004)	0.00039*** (0.0001)	0.00051*** (0.0001)
CDD	-0.00008 (0.0003)	-0.00016 (0.0002)	-0.00007 (0.0001)
DDd	-0.00028 (0.0002)	0.00043*** (0.0001)	0.00017 (0.0001)
Δ Log Oil Price	0.106 (0.157)	-0.123 (0.115)	-0.150* (0.088)
Δ Log LNG Price		0.694** (0.056)	0.590*** (0.058)
Seasonal Lag		0.69370*** (0.05569)	0.59027*** (0.05819)
Observations	167	156	156
R ²	0.905	0.955	
Adjusted R ²	0.902	0.953	
Log Likelihood			158.531
Akaike Inf. Crit.			-297.063
Bayesian Inf. Crit.			-266.564
Residual Std. Error	0.171 (df = 161) 308.309*** (df = 5;	0.120 (df = 149) 523.714*** (df = 6;	0.13 (df=149)
F Statistic	161)	149)	
	*p<.10 **p<.05 ***p<0.01		

Note: Models for the firm contract quantity consumed are presented in each column. The GLS regression specifies a ARMA(2,0) errors based on minimizing AIC from alternate lag specifications. The results are used to estimate the leftover natural gas capacity for the electric sector. The simulation uses these fitted firm quantity estimates to create the capacity left for the electric sector.

TABLE 7
Equilibrium Natural Gas Price Under Increased Capacity Scenarios
New England Electric Sector

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Mild Winter												
DDd	9	-112	46	-147	-95	-27	3	-7	-10	48	126	-31
Current	4.84	5.16	6.64	7.66	11.07	7.96	7.1	5.76	5.25	5.37	5.41	4.81
+450MMcf/d	4.6	4.89	6.19	7.04	9.91	7.25	6.51	5.4	4.97	5.1	5.15	4.58
Difference	-0.24	-0.27	-0.45	-0.62	-1.16	-0.71	-0.59	-0.36	-0.28	-0.27	-0.26	-0.23
Median Winter												
DDd	19	24	31	-30	171	-31	26	-62	-15	10	-43	58
Current	4.88	5.7	6.56	8.63	16.15	7.92	7.26	5.52	5.22	5.16	4.65	5.19
+450MMcf/d	4.64	5.39	6.12	7.87	13.75	7.22	6.66	5.18	4.94	4.9	4.42	4.94
Difference	-0.24	-0.31	-0.44	-0.76	-2.40	-0.70	-0.60	-0.34	-0.28	-0.26	-0.23	-0.25
Severe Winter												
DDd	-69	17	41	-48	228	-41	-38	-50	-37	4	-20	-14
Current	4.57	5.67	6.61	8.47	17.9	7.94	6.82	5.57	5.14	5.16	4.77	4.89
+450MMcf/d	4.34	5.37	6.17	7.73	14.94	7.23	6.27	5.22	4.87	4.91	4.54	4.65
Difference	-0.23	-0.3	-0.44	-0.74	-2.96	-0.71	-0.55	-0.35	-0.27	-0.25	-0.23	-0.24

Note: Prices are differences in \$/Mcf. Scenarios based on the regression results from column six of Table 3. The winter scenarios present a range of severities of winters. Price is highest in January. The effect of increased capacity is largest in more severe winters.

TABLE 8
Equilibrium Natural Gas Quantities Under Increased Capacity Scenarios
New England Electric Sector

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Mild Winter												
DDd	9	-112	46	-147	-95	-27	3	-7	-10	48	126	-31
Current	32799	29920	27152	27169	24050	25628	26548	28995	30482	37221	47576	36069
+450MMcf/d	33546	30643	27995	28198	25246	26698	27570	29822	31218	38056	48631	36868
Difference	747	723	844	1029	1196	1071	1022	828	736	835	1054	799
Median Winter												
DDd	19	24	31	-30	171	-31	26	-62	-15	10	-43	58
Current	33071	29180	27248	26206	21134	25661	26364	29641	29861	33384	36885	39872
+450MMcf/d	33825	29904	28090	27288	22683	26730	27396	30470	30583	34136	37702	40758
Difference	753	724	841	1082	1549	1069	1032	829	722	752	817	885
Severe Winter												
DDd	-69	17	41	-48	228	-41	-38	-50	-37	4	-20	-14
Current	32212	29218	27184	26358	20359	25651	26872	29307	30153	34775	39560	37108
+450MMcf/d	32940	29942	28027	27431	22042	26733	27878	30132	30878	35555	40434	37930
Difference	728	724	843	1073	1683	1082	1006	825	726	780	875	822

Note: Consumption and differences is in MMcf. Scenarios based on the regression results from column six of Table 5. The winter scenarios present a range of severities of winters.

TABLE 9
New England Market for Natural Gas for the Electric Sector
3SLS SEM Estimates

NG Demand Equation		NG Supply Equation		Firm Quantity		Working Gas	
NG quantity consumed	-0.567** (0.230)	NG for electric price	-0.121*** (0.032)	HDD	44.967*** (2.076)	Firm Quantity	-0.000069*** (0.00001)
HDD	.00018** (-0.0001)	Log NG Capacity	0.713** (0.290)	CDD	40.174*** (5.868)	Quantity Electric	.000016** (-0.00001)
CDD	.0011* (-0.0005)	HDD	0.0004** (0.0002)	DDd	-14.350*** (4.674)	HDD	0.003*** (0.0003)
DDd	0.00072** (-0.0002)	CDD	.0018*** (0.0002)	Δ Log Oil Price	-566.816 (2,617.112)	CDD	0.002*** (0.001)
Δ Log Oil Price	0.057 (0.190)	DDd	-.000061 (0.0001)	Δ Log LNG Price	13.463 (103.250)	DDd	-0.001*** (0.0003)
Working Gas	0.221*** (0.048)	Δ Log Oil Price	-0.112 (0.111)	Working Gas	9,539.078*** (1,283.210)	R-squared	0.60
R-squared	0.80	Δ Log LNG Price	0.006 (0.004)	R-squared	0.97	DF	137
N	144	Working Gas	-0.071 (0.050)	N	144		
DF	133			DF	134		
						System	
		R-squared	0.73			R-Squared	0.97
		N	144			N	576
		DF	132			DF	536

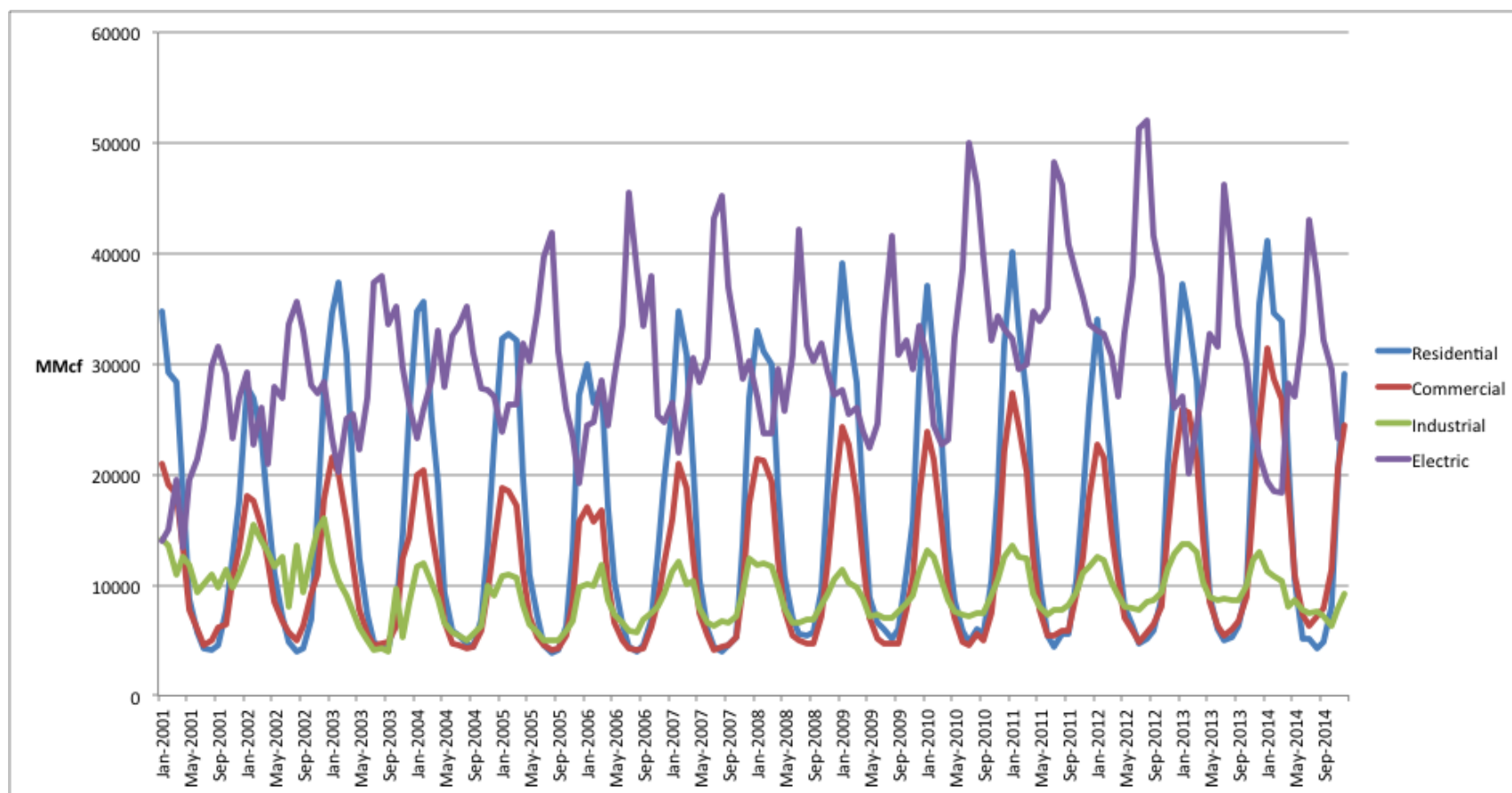
Note: Each column presents the results of the 3SLS estimates of each structural equation. Time trends, seasonal and one period lags are included of each endogenous variable. Standard errors are in parentheses. *p<.10 **p<.05 ***p<0.01

TABLE 10
New England Market for Natural Gas for the Electric Sector
2SLS SEM Estimates

NG Demand Equation		NG Supply Equation		Firm Quantity		Working Gas	
NG quantity consumed	-0.419*	NG for electric price	-0.113***	HDD	38.384***	Firm Quantity	-0.00005***
	(0.236)		(0.034)		(2.728)		(0.00001)
HDD	0.0002***	Log NG Capacity	0.602**	CDD	38.037***	Quantity Electric	0.00003***
	(0.0001)		(0.305)		(5.942)		(0.00001)
CDD	0.001	HDD	0.0004*	DDd	-0.672	HDD	0.002***
	(0.001)		(0.0002)		(5.277)		(0.0004)
DDd	0.001***	CDD	0.002***	Δ Log Oil Price	-285.735	CDD	0.0005
	(0.0002)		(0.0003)		(3,563.617)		(0.001)
Δ Log Oil Price	0.121	DDd	-0.0001	Δ Log LNG Price	75.274	DDd	-0.001***
	(0.193)		(0.0001)		(141.149)		(0.0003)
Working Gas	0.238***	Δ Log Oil Price	-0.124	Working Gas	-966.776	R-squared	0.75
	(0.049)		(0.111)		(1,771.276)	N	144
R-squared	0.80	Δ Log LNG Price	0.009*	R-squared	0.97	DF	137
			(0.005)				
N	144			N	144		
DF	133	Working Gas	-0.073	DF	134		
			(0.052)				
						System	
		R-squared	0.77			R-Squared	0.97
		N	144			N	576
		DF	132			DF	536

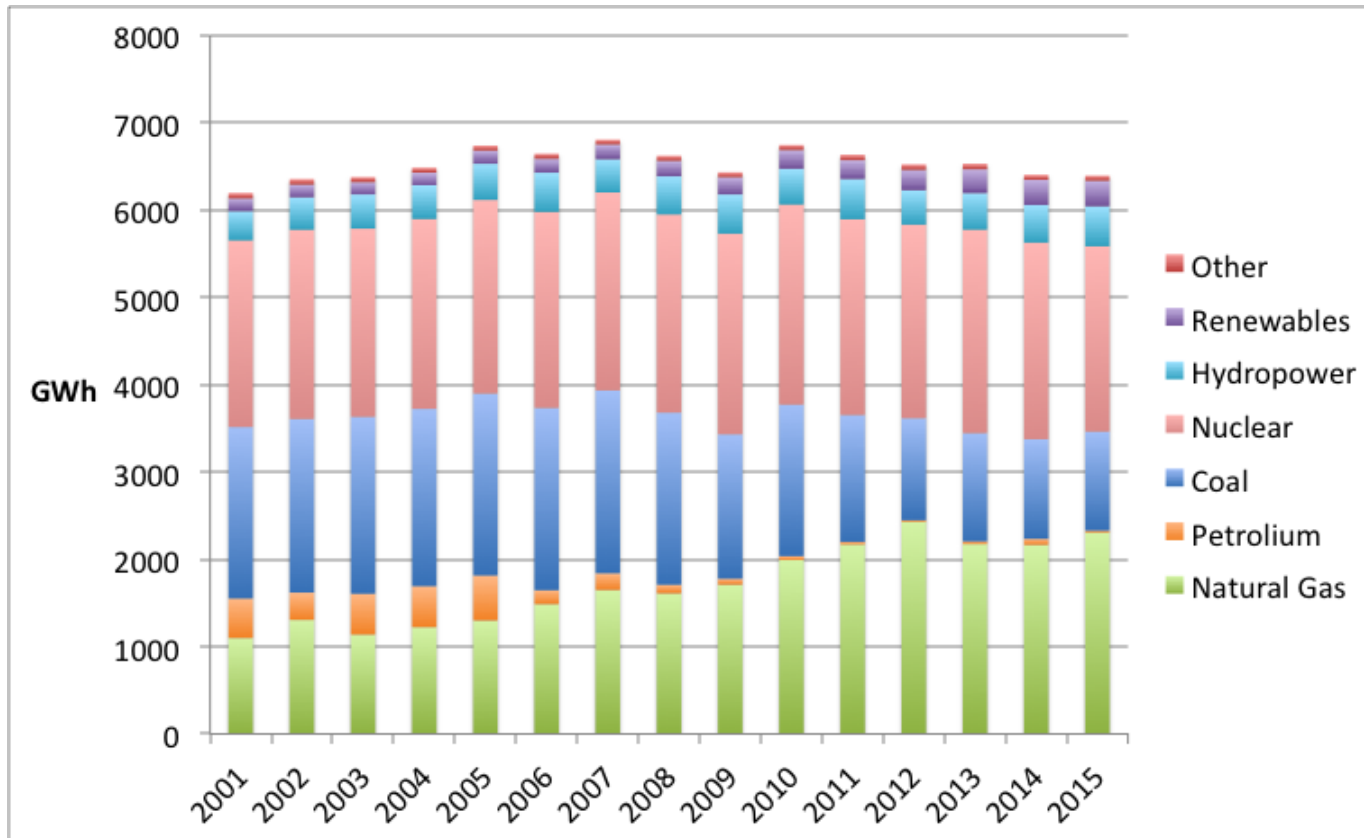
Note: Each column presents the results of the 2SLS estimates of each structural equation. Time trends, seasonal and one period lags are included of each endogenous variable. Standard errors are in parentheses. *p<.10 **p<.05 ***p<.01

FIGURE 1
Natural Gas Consumption by Sector
New England



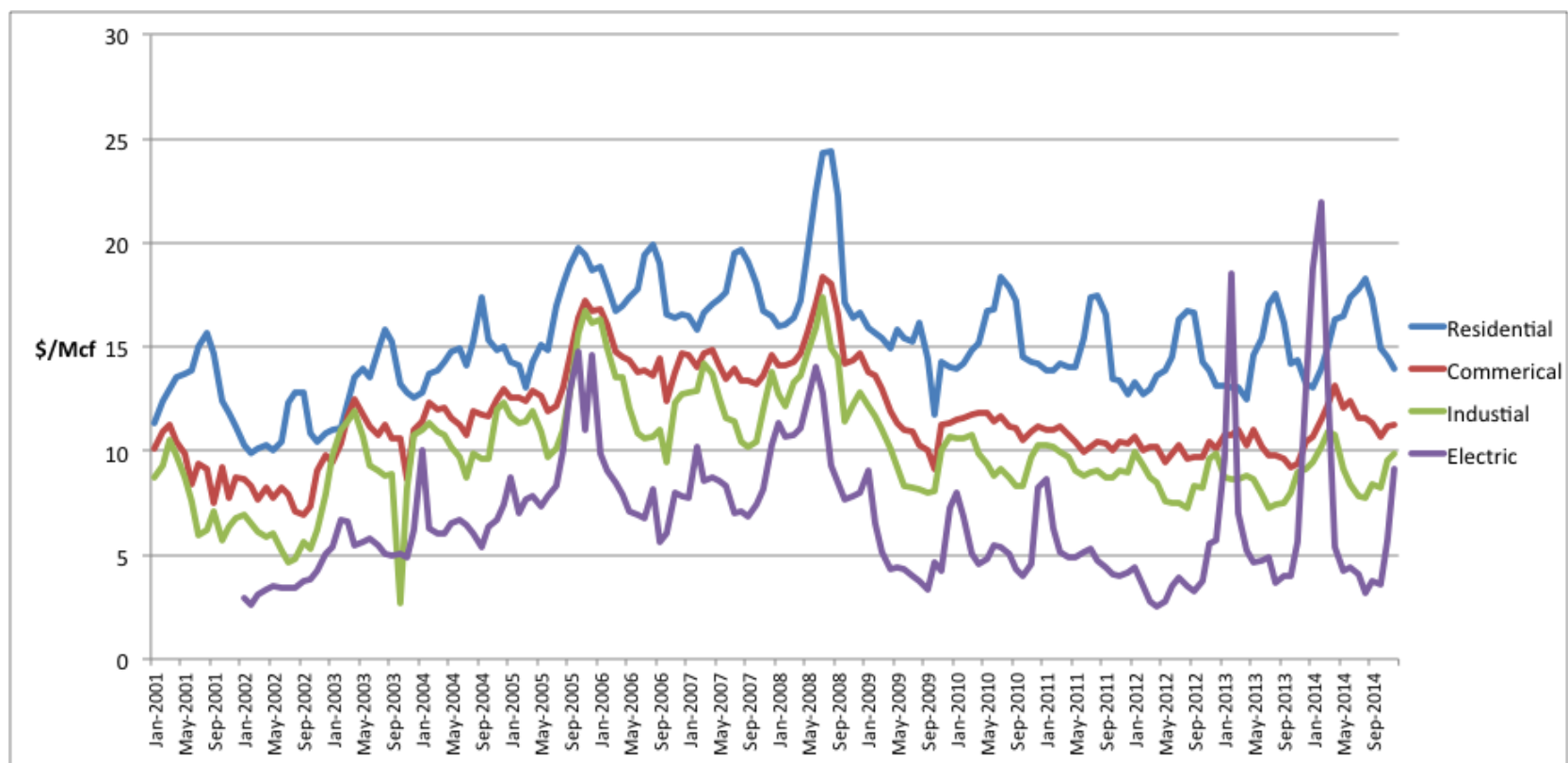
Note: Data is from the EIA. Residential, commercial and industrial uses increase in the winter due to heating requirements, while electric energy uses peak in the summer due to cooling demand.

FIGURE 2
Electric Energy Generation by Fuel Source
New England



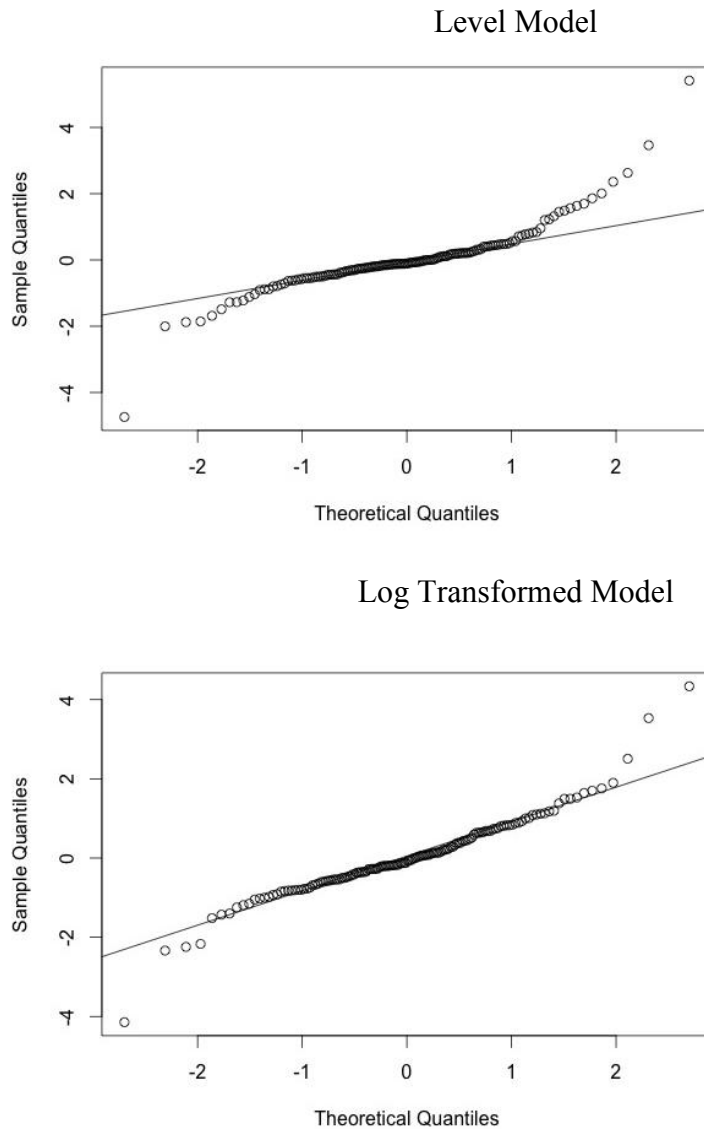
Note: Data is from ISO-NE and represents the total GWh from each source for the year. Natural gas generators have replaced retiring coal plants. Total consumption has remained mostly constant.

FIGURE 3
Natural Gas Price by Consuming Sector
New England



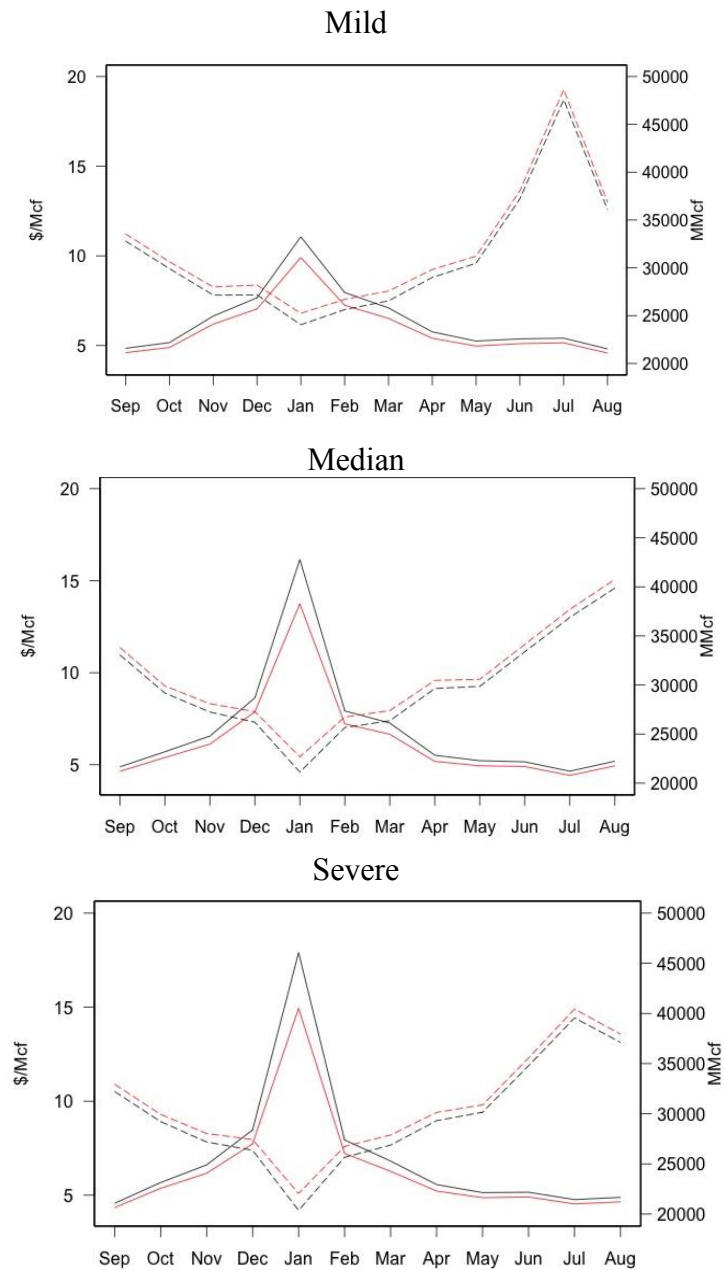
Note: The price data is from surveys conducted by the EIA and accessed through the EIA's data portal. Each sector's price is the per unit price of natural gas including delivery. Residential, industrial and commercial prices follow a different path based on the structure of the local utilities. The electric sector operates under interruptible contracts, which leads to increases in times of pipeline constraint.

FIGURE 4
Log Transformation Q-Q Plot Test
Preferred model



Note- This figure shows the quintile plot of the residuals from the preferred price model, column 6 in Table 3, when specifications with the dependent and independent variables are in levels compared to in logs. The temperature variables remain in levels in both models. The log transformations create errors that are closer to normally distributed, the strait lines in the plots.

FIGURE 5
Simulation of Equilibrium Price and Quantity
New England Electric Sector



Note: Price and quantity projections of an increase in 450 MMcf/d of pipeline capacity. Each plot represents a different winter weather scenario. Price is in solid lines and quantity in dashed lines for both the current (black) and future (red) capacity scenarios.

MANUSCRIPT 2 APPENDIX

Derivation of Equation 1

The area of a circle representing the irrigated agriculture acreage above the aquifer.

$$A = \pi u^2$$

With a fixed initial aquifer area and maximum depth, we can relate geometry of the aquifer to the radius of the irrigated acreage

$$r = \frac{x}{\tan(\theta)}$$

Using these two, we get irrigated area as a function of the groundwater height and geometry of the aquifer.

$$I = \pi \left(\frac{x}{\tan(\theta)} \right)^2$$

With the initial conditional known, we can solve for initial radius and $\tan(\theta)$:

$$r_0 = \frac{x_0}{\tan(\theta)}$$

$$\tan(\theta) = \frac{x_0}{u_0}$$

With these we can come to equation (1) in the paper by making at a percent of the initial aquifer area:

$$\gamma_t(x_t) = \frac{\pi \left(\frac{x_t}{x_0/u} \right)^2}{A}$$

Discussion of Equation 8

Here we want to illuminate the model dynamics of including a cone and uniform spatial depletion of the groundwater resource. The speed of depletion is important in the model and is determined by the irrigated area and thus a function of the height of groundwater; therefore we compare the speed of height change in our model to the speed of height change in a bathtub model.

Proposition 1: $\dot{x}_{sp} < \dot{x}_{bt}$ given the same demand for water and a decreasing path of groundwater height. The change in groundwater height is larger in the bathtub model than in the spatial depletion model.

We start with defining per acre water demand constant to give a fair comparison between models

$$w = g + kP \quad (1a)$$

where w is the per acre water demand in the bathtub and spatial depletion model at a given height defined by the price of water P . The total amount of water extracted is a function of the height of the water as is the change in groundwater height. The total water extracted decreases in the spatial model because there are less irrigated acres as the height of groundwater decreases. This is realized through the spatial depletion function $\gamma_t(x_t)$ or equation (1) in the main text. The rate at which groundwater height changes, the equation of flow, is also a function of $\gamma_t(x_t)$. \dot{x}_{sp} refers to the spatial depletion model and \dot{x}_{bt} refers to the bathtub model.

$$\dot{x}_{sp} = \frac{R - (1 - \alpha)A\gamma_t(x_t)w_t}{\gamma_t(x_t)AS} \quad (2a)$$

$$\dot{x}_{bt} = \frac{R - (1 - \alpha)Aw_t}{AS} \quad (3a)$$

and

$$\frac{\partial \gamma_t(x_t)}{\partial x} = \frac{2\pi \left(\frac{u}{x_0}\right)^2 x_t}{A} > 0 \quad (4a)$$

Equation (4a) describes how the percent of irrigated area of the spatial depletion model changes. Increase in height cause an increase in irrigable area, and a decrease in height causes a decrease in irrigable area at this rate described in in (4a). Taking a difference in equation (2a) and (3a) we find

$$\dot{x}_{sp} - \dot{x}_{bt} = \frac{R}{\gamma_t(x_t)AS} - \frac{R}{AS} \quad (5a)$$

$$\lim_{\gamma \rightarrow 0} (\dot{x}_{sp} - \dot{x}_{bt}) = \infty \quad (6a)$$

$$\lim_{\gamma \rightarrow 1} (\dot{x}_{sp} - \dot{x}_{bt}) = 0 \quad (7a)$$

Equations (6a) and (7a) state that as the aquifer is depleted the change in height is less in the spatial the depletion model at the rate given in equation (4a) and because the spatial component becomes smaller and is in the denominator of equation (2a) the recharge acts as a buffer to loss in height in the spatial depletion model. As the depletion parameter, γ , gets closer to one the difference converges to zero. As x decreases so does γ , which drives the differential to be positive.

Proposition 2: $\dot{x}_{sp} = \dot{x}_{bt}$ *when there is compaction of the aquifer which reduces storage in the spatial depletion model under the same per acre water demand.*

A special case of equation (2a) is when total recharge decreases as the height declines, this could be the case when subsidence occurs which shuts out storage to the aquifer (Scanlon et al. 2012). If recharge is scaled to the remaining irrigable acreage to reflect this subsidence and loss in storage which we define as

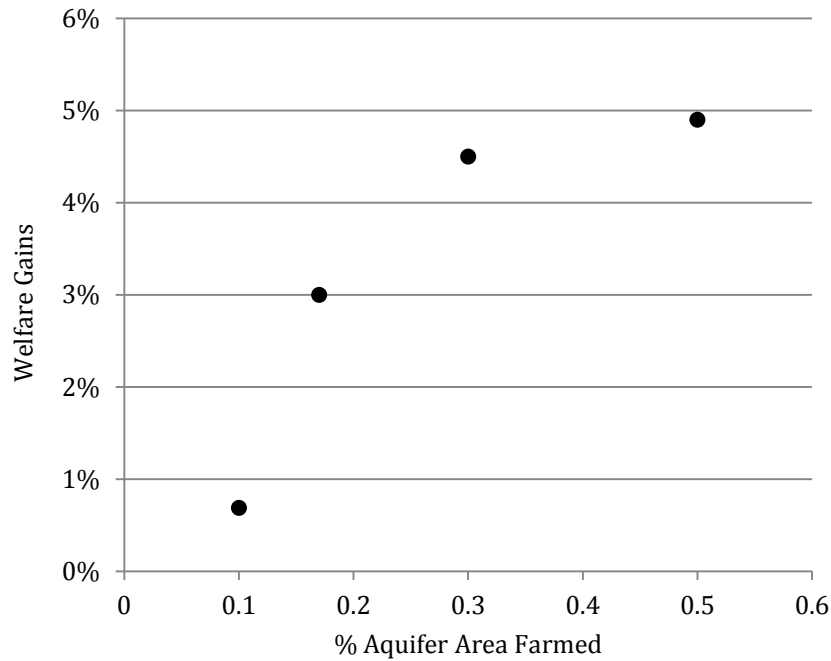
$$R = recharge * A\gamma_t(x_t) \quad (8a)$$

The equation flows are equal under the same demand curve.

$$\dot{x}_{sp} = \dot{x}_{bt} = \frac{recharge * A - (1 - \alpha)Aw_t}{AS} \quad (9a)$$

Scanlon, Bridget R., et al. "Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley." *Proceedings of the national academy of sciences* 109.24 (2012): 9320-9325.

APPENDIX FIGURE 1
Welfare Gains to Management as a function of farming intensity
Stochastic Scenario



NOTE- Welfare gains from management are sensitive to the assumption of farming intensity which defines demand for groundwater and its marginal value. As farming intensity is increased so to are the estimated gains to management.

APPENDIX TABLE 1
Welfare Gains From Groundwater Management
Stochastic Rainfall Scenario
Total Discounted Profit (Billion \$)

	Perfect Competition	Optimal Policy	Difference	% Gain
Stochastic - Spatial	\$ 6.589	\$ 6.785	\$.196	2.97
Depletion	(.0937)	(.0938)	(.006)	(.04)
Stochastic - No Backstop	\$ 6.021	\$ 6.335	\$.313	5.20
	(.0623)	(.0670)	(.006)	(.06)
Stochastic - Bathtub Model	\$ 8.297	\$ 8.302	\$.0052	0.06
	(.0063)	(.0054)	(.0015)	(.02)

NOTE- Standard errors of the stochastic figures from 100 iterations through rainfall realizations are in parentheses. The deterministic scenario assumes average annual rainfall each year. Stochastic assumes i.i.d. random draws from high, medium, low rainfall state based on empirical probabilities. The bathtub model assumes no loss in irrigated land as groundwater levels fall. The no backstop scenario assumes that as irrigated land falls, no farming at all occurs on acreage that no longer has access to the aquifer.

ARIMA Reduced Form Model

The I in ARIMAX stands for the level of integration, or how many times the dependent variable must be differenced to be stationary. Using the stationary series, having first differenced the non-stationary variables, an ARMAX of the following form is estimated.

$$y_t = c + \sum_{p=1}^p \beta_p y_{t-p} + \sum_{q=1}^q \theta_q e_{t-q} + X_t \gamma + e_t$$

Where y_{t-p} is the lagged, p period, dependent variable with coefficients β_p . Seasonal lags enter in the same manner, but lagged 12 periods on our model. These terms handle the autoregressive, AR_p , process. θ_q are coefficients on the moving average, MA_q , process on previous residuals of a lag length q . X is a matrix of exogenous regressors with coefficient estimates γ . The appropriate length of the lag is a choice of the modeler, but can be guided by evaluating the autocorrelation correlation function (ACF) and partial autocorrelation function (PACF). From there it is common to use measures of AIC and BIC can determine the most appropriate lag lengths.

I estimate the reduced form equations for both price and quantity based on the criteria described above, which turned out to be the lag structures found in the OLS residuals and specified in the GLS regressions from Table 3 and 5. The price equation is of the form $ARIMAX(1,0,0)(1,0,0)_{12}$ and the quantity equation of the form $ARIMAX(2,0,0)(1,0,0)_{12}$, where $ARIMAX(p,d,q)(P,D,Q)_m$. The second set of parentheses specifies the seasonal terms. M is the seasonal period lag. The results are presented in Table 8.

The results are similar to that found in the GLS regressions with an increase in leftover capacity effecting prices negatively. The coefficient is larger in magnitude, implying a 1% change in capacity leading to a .6% change in price. Degree day anomalies are significant and of similar magnitude. The quantity equation similarly shows a larger effect of the leftover capacity on the equilibrium quantity consumed.

The results confirm the results of the GLS regressions. A benefit of the ARIMAX models is that predictions can be made that take into account the time series of the error structure, as opposed the GLS results, which use the unbiased estimates of the coefficients to estimate values, but does not take into account the recent error term innovations in making step ahead forecasts. Therefore, once values for the exogenous regressors are provided, future values can be assessed under both current and increased capacity. Similarly to the policy simulations above, first I estimate the firm demand, using an ARIMAX model, calculate the remaining capacity and estimate the price and quantity for electric energy generation equations. I estimate 2 years ahead, with the time series of the severe winter scenario. The results are in Figure 6.

APPENDIX FIGURE 1
Natural Gas Price and Quantity for Electric Energy Generation
ARMAX Models

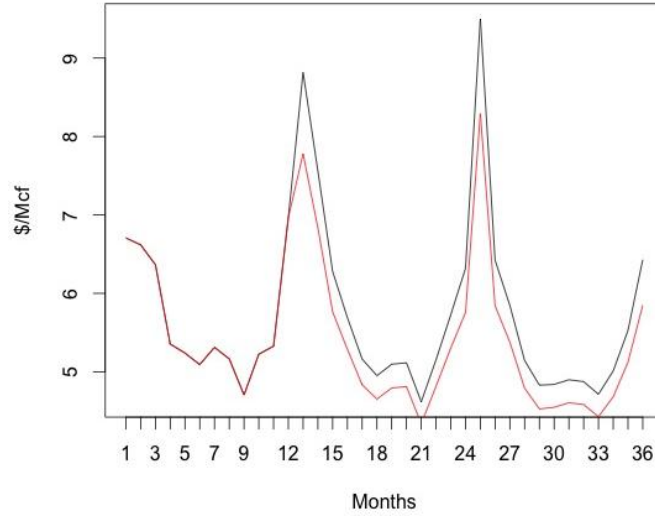
<i>Dependent variable:</i>		
	NG Electric Price	NG Electric Quantity
Model Type	ARIMAX(1,0,0)(1,0,0) ₁₂	ARIMAX(2,0,0)(1,0,0) ₁₂
Log NG Capacity	-0.60164 ^{***} (0.19503)	0.29561 ^{**} (0.12055)
HDD	0.00004 (0.00021)	0.00006 (0.00012)
CDD	0.00031 (0.00035)	0.00161 ^{***} (0.00024)
DDd	0.00051 ^{**} (0.00020)	-0.00006 (0.00015)
Δ Log Oil Price	0.04783 (0.15305)	-0.19237 [*] (0.10412)
Δ LNG Price	0.00524 (0.00427)	0.00584 [*] (0.00306)
AR1 Lag	0.88330 ^{***} (0.03803)	0.63403 ^{***} (0.06562)
Seasonal Lag	0.28594 ^{***} (0.08417)	0.30880 ^{***} (0.09055)
Log Likelihood	48.69	129.07
AIC	-76.07	-238.14
BIC	-45.44	-206.21

* ** *** p p p<0.01

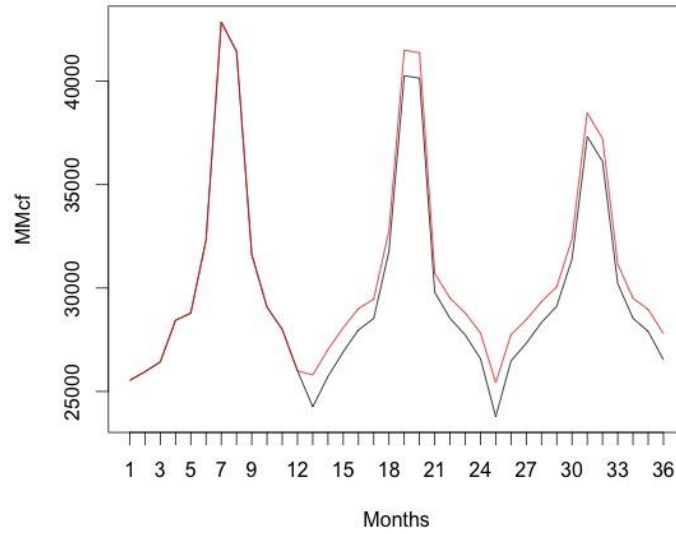
Note: Reduced from equation of the natural gas price and quantity for the electric energy generating sector are presented. The lag structure is based on minimizing AIC with the inclusion of the exogenous covariates and a 12-month seasonal lag.

APPENDIX FIGURE 2
Natural Gas Prices for Electric Energy Generation
Equilibrium Price and Quantity
ARIMAX Models

Price



Quantity



Note: Price and Quantity projections from the ARIMAX model in Table 8. Months count ahead from January, 2015 to December 2017. The red line represents an increase of 450 MMcf/d of pipeline capacity added at the start of 2016, or 12 months ahead in the figure. The increased capacity leads to lower prices and higher quantity consumed.